Technical Report 1145

Validation of Measures Designed to Maximize 21st-Century Army NCO Performance

Deirdre J. Knapp and Rodney A. McCloy (Eds.)
Human Resources Research Organization

Tonia S. Heffner(Ed.)
U.S. Army Research Institute

May 2004

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute for the Behavioral and Social Sciences

A Directorate of the U.S. Army Human Resources Command

ZITA M. SIMUTIS
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical Review by

Peter Greenston, U.S. Army Research Institute
D. Bruce Bell, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-PO, 5001 Eisenhower Ave., Alexandria, VA 22304-4841.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
May 2004	Interim	February 2001 – June 2002

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Validation of Measures Designed to Maximize 21st-Century Army NCO Performance	DASW01-98-D-0047/DO #15
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Deirdre J. Knapp & Rodney A. McCloy (Eds.) (Human Resources Research Organization), and Tonia S. Heffner (Ed.) (U.S. Army Research Institute)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 6900B
	5e. WORK UNIT NUMBER C05

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Human Resources Research Organization 66 Canal Center Plaza, Suite 400 Alexandria, VA 22314	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. MONITOR ACRONYM
U. S. Army Research Institute for the Behavioral & Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22304-4841	ARI
	11. MONITOR REPORT NUMBER Technical Report 1145

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Tonia Heffner, Contracting Officer's Representative

14. ABSTRACT *(Maximum 200 words)*: The NCO21 research program was undertaken to help the U.S. Army plan for the impact of future demands on the noncommissioned officer (NCO) corps. The performance requirements and associated knowledges, skills, and aptitudes (KSAs) expected of future successful NCOs were used as a basis for developing tools that could be incorporated into an NCO performance management system geared to 21st-century job demands. This report documents the concurrent criterion-related validation of the predictor measures.

The predictor measures include the Armed Services Vocational Aptitude Battery (ASVAB), Assessment of Individual Motivation (AIM), and Biographical Information Questionnaire (BIQ), which are already used in the Army for other purposes. A written Situational Judgment Test (SJT), the Experience and Activities Record (ExAct), Personnel File Form (PFF21), and a semi-structured interview were developed for this project. Two types of rating scale instruments were developed for gathering criterion data. The Observed Performance Rating Scales ask supervisors to rate how well Soldiers perform in their current jobs and the Expected Future Performance Rating Scales have them predict how their Soldiers would perform in future conditions.

All of the predictors yielded one or more scores that were significantly correlated with the performance ratings. The SJT, interview, and some scores from the AIM and BIQ showed the most incremental validity over the current system. In this concurrent validation, the predictors were more highly correlated with performance at the E5 level compared to the E6 level.

15. SUBJECT TERMS

Behavioral and social science	Personnel	Promotion
Selection and Classification	Manpower	

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	224	Tonia Heffner (703) 617-8557

Standard Form 298

Technical Report 1145

Validation of Measures Designed to Maximize 21st-Century Army NCO Performance

Deirdre J. Knapp and Rodney A. McCloy (Eds.)
Human Resources Research Organization

Tonia S. Heffner (Ed.)
U.S. Army Research Institute

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22304-4841

May 2004

Army Project Number
2O262785A790

Personnel Systems and
Performance Technology

FOREWORD

This project, entitled "NCO21: 21st-century Noncommissioned Officer Requirements," is being conducted by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) under the sponsorship of the Army G-1. The goal of NCO21 is to conduct an analysis of future conditions and future job demands in order to identify critical performance predictors--knowledges, skills, and aptitudes (KSAs)--that may eventually be used to select and grow future noncommissioned officers (NCOs). This project has been divided into three phases. Completion of the first two phases was documented in earlier reports. Phase I was the development of a detailed research plan for identifying characteristics required of future NCOs. In Phase II, the methodological steps of the Phase I research plan were executed. Anticipated job requirements of 21st-century NCOs (for the years 2000 through 2025) were forecasted and the most important KSAs needed for success in Army jobs were estimated.

Phase III involves the remainder of the project activities, including development and validation of KSA measures. This report documents the second stage of Phase III, which involved the collection and analysis of criterion-related validation data. The information presented in this report was briefed to the Chief, Enlisted Division, Directorate of Military Personnel Management, Deputy Chief of Staff for Personnel (DCSPER) and the DCSPER Sergeant Major on 13 August 2001. It was briefed to U.S. Army Training and Doctrine Command (TRADOC) representatives on 11 October 2001 and briefed to the Commanding General, U.S. Total Army Personnel Command (PERSCOM) on 29 July 2002. Uses of the tools developed in this effort will be determined in discussions with ODCSPER and TRADOC representatives based on the findings obtained from the Phase III validation.

The goal of the Selection and Assignment Research Unit of ARI is to conduct research, studies, and analysis on the measurement of aptitudes and performance of individuals to improve the Army's selection and classification, promotion, and reassignment of officers and enlisted Soldiers. This research will provide the foundation for recommended improved promotion and development procedures for enlisted personnel.

STEPHEN L. GOLDBERG
Acting Technical Director

Acknowledgements

Many people contributed to the work documented in this report. Those not listed as authors include Dr. Michael G. Rumsey and Dr. Trueman Tremble from the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) and Dr. William Strickland from the Human Resources Research Organization (HumRRO). Data collection staff included Lori Boyd, Jennifer Burnfield, Doug Brown, Charlotte Campbell, Roy Campbell, Mike Cobb, Ani DiFazio, Laura Ford, Chelsey Hibbard, Art Paddock, Chris Sager, and Gordon Waugh (HumRRO) and Tonia Heffner, Mike Rumsey, and Shawn Bergman (ARI). We gratefully acknowledge everyone's ideas, feedback, and efforts related to this work. We also appreciate the tremendous secretarial support provided by Ms. LaVonda Murray of HumRRO.

VALIDATION OF MEASURES DESIGNED TO MAXIMIZE 21ST-CENTURY ARMY NCO PERFORMANCE

EXECUTIVE SUMMARY

Research Requirement

The NCO21 research program was undertaken to help the U.S. Army plan for the impact of future demands on the noncommissioned officer (NCO) corps. When the NCO21 research program began, a great deal of effort was being devoted to analyzing national and global trends (e.g., more complex technology with increasingly sophisticated capabilities, demographic changes) that would presumably affect the U.S. military in terms of its missions, organizational structure and technology, strategies and tactics, and personnel systems. But these analyses and forecasts were not available in any consolidated form. Indeed, there was (and still is) considerable variation in the prognostications being made. Moreover, little had been done to look at the implications of expected future changes for the performance requirements of individual Soldiers. The purpose of the first stage of this research program, then, was to (a) identify and review the available information on predictions and plans related to the Army's future and (b) attempt to abstract from these a reasonable idea of what performance expectations would be imposed on NCOs of the future. In subsequent stages of the research program, these expectations have been used to develop procedures and methods that could be incorporated into the NCO performance management system in an effort to make the NCO corps better prepared to handle 21st-century job demands. Specifically, predictor and criterion (job performance) measures were designed and developed for use in a concurrent criterion-related validation effort. This report describes the validation data collection and analysis work. It is primarily targeted toward a technical audience interested in the psychometric characteristics and quality of the measures.

Procedure

There were seven predictor measures to be validated. Three measures—the Armed Services Vocational Aptitude Battery (ASVAB), Assessment of Individual Motivation (AIM), and Biographical Information Questionnaire (BIQ)—are operational tests (in whole or in part) already used in the Army for other purposes. Experimental versions of the AIM and BIQ were prepared for use in the present research. Four measures—a written Situational Judgment Test (SJT) (and its close cousin, the SJT-X), the Experience and Activities Record (ExAct), the Personnel File Form (used to compute a Promotion Point Worksheet score that simulates the current promotion system), and a semi-structured interview—were developed for this project.

The predictor measures were validated by examining how well they predicted job performance as assessed using two types of supervisor rating scale instruments. The Observed Performance Rating Scales ask supervisors to rate Soldiers on how well they perform in their current jobs. The Expected Future Performance Rating Scales ask supervisors to predict how their Soldiers would perform in specific sets of conditions expected to be characteristic of future Army requirements.

Predictor data were collected from roughly 1,900 Soldiers in pay grades E4 though E6. Performance ratings were collected for about 70% of the E5 and E6 Soldiers, so they constituted the primary validation sample.

Findings

The results of the validation analyses were very promising. All of the predictor instruments yielded one or more scores that were significantly correlated with performance, both current and future. Even when examining incremental validity over the current system, most instruments performed well. The SJT, interview, and scores from the AIM and BIQ showed the highest incremental validity. Complicating the analyses and subsequent conclusions was the finding that the empirical results varied across pay grade and career management field (CMF). Despite extensive analyses to identify artifactual source(s) of these differences (e.g., range restriction), none were found.

Utilization of Findings

The findings reported here will be the basis for recommendations made to the Army about the possible implementation of the NCO21 measures – the subject of a companion report. Although the evidence supporting implementation of several of the NCO21 measures is quite positive, it is based on a concurrent validation sample in a research setting. Additional research using a longitudinal design in an operational setting is recommended to support the assignment of promotion points in the Army's semi-centralized NCO promotion system based on any of these new measures.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Appendixes

List of Tables

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

List of Figures

VALIDATION OF MEASURES DESIGNED TO MAXIMIZE 21ST-CENTURY ARMY NCO PERFORMANCE

CHAPTER 1: INTRODUCTION

Deirdre J. Knapp and John P. Campbell
HumRRO

This report describes the concurrent criterion-related validation of a set of experimental noncommissioned officer (NCO) promotion tools, part of a multi-phased research program sponsored by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). The report is targeted primarily toward a technical audience interested in the psychometric characteristics and estimated validity of the measures. Those readers interested in more detail on the development and field testing of the measures should see Knapp et al. (2002).

Background

Overview of the NCO21 Research Program

The NCO21 research program was undertaken to help the U.S. Army understand and plan for the impact of future performance demands on the future NCO performance management system. When the research program began, much effort was being devoted to analyzing national and global trends (e.g., more complex technology with increasingly sophisticated capabilities, demographic changes) that would presumably affect the U.S. military in terms of its missions, organizational structure and technology, strategies and tactics, and personnel systems. But these analyses and forecasts were not available in any consolidated form. Indeed, there was (and still is) considerable variation in the prognostications being made. Moreover, very little had been done to look at the implications of expected future changes for the performance requirements of individual Soldiers. The purpose of the first stage of this research program, then, was to (a) identify and review the available information on predictions and plans related to the Army's future and (b) attempt to abstract from these a reasonable idea of what performance expectations would be imposed on NCOs of the future. In subsequent stages of the research program, these expectations have been used to develop procedures and methods that could be incorporated into the NCO performance management system in an effort to make the NCO corps better prepared to handle 21st-century job demands.

Following some preliminary efforts conducted by ARI staff, the NCO21 research program was divided into three phases, each of which has been supported through a contract to the Human Resources Research Organization (HumRRO):

- Phase I: Develop a method to identify future job requirements (J. Campbell, Walker, & Knapp, 1998).
- Phase II: Forecast future NCO performance requirements and the individual characteristics necessary to meet those requirements (Ford, Knapp, J. Campbell, R. Campbell, & Walker, 2000).

- Phase III: Develop measures of the relevant variables (Knapp et al., 2002), conduct validation research to estimate their usefulness, and make recommendations for potential changes to the NCO promotion system. (The validation was completed in 2001 and is the subject of this report; recommendations are documented in a separate report – see Knapp, Heffner, & R. Campbell, 2003).

NCO21 Job Analysis (Phases I and II)

The Phase II final report documents the collection and integration of future projections (Ford et al., 2000). It also describes the construction of baseline (1990s) information about NCO requirements— in terms of both performance requirements (e.g., motivating and leading others) and the knowledges, skills, and aptitudes (KSAs) required for successful job performance (e.g., general cognitive aptitude, conscientiousness). The baseline requirements were then updated based on an analysis of conditions in two future eras (the period 2000-2010 and the period 2010-2025). Two expert panels (one comprising Army subject matter experts [SMEs] and another comprising personnel psychologists) used this information to judge the relative importance of KSAs for the different time periods. Phase II thus generated the products listed below.

- Descriptions of the forecasted job demands for two future eras (2000-2010, 2010-2025)
- Lists of performance requirements for three eras (1990s baseline, 2000-2010, 2010-2025)
- Prioritized lists of KSAs for all three eras

Because of differences in NCO requirements across ranks, the baseline and 2000-2010 era KSA priority rankings were determined separately by NCO level: junior (E4/E5), mid-level (E6/E7), and senior (E8/E9). The 2010-2025 era was forecasted to incorporate the Army envisioned for the 2000-2010 era supplemented by a "Battleforce" component comprising more experienced and specialized Soldiers. Therefore, the 2010-2025 era KSAs were prioritized simply for Battleforce NCOs, irrespective of rank.

When the NCO21 job analysis work was conducted, the Army used different terms to characterize its future (e.g., the Army After Next). Since then, the language has changed (we now speak of the Objective Force), the planning time horizon has been extended beyond the mid-point of the century, and some future plans have become more fully realized and/or articulated. Despite these changes, there have not been significant changes in direction that invalidate the future-oriented job analysis work conducted 3 years ago. That is, were we to conduct the job analysis again today, we would not expect to obtain substantially different results.

Instrument Development, Validation, and Recommendations (Phase III)

Whereas Phase II focused on Soldier requirements across all NCO levels (shown in Table 1.1), the focus in Phase III narrowed to the semi-centralized NCO promotion system. This system covers promotions from grade E4 to E5 and from grade E5 to E6. It was necessary to narrow the focus because of the inordinate resources required to develop and validate measures suitable across all NCO ranks. The semi-centralized promotion system, however, covers more than 70% of the Army NCO corps, so improving this system would have a substantial impact.

Table 1.1. U.S. Army NCO Pay Grades and Ranks

Pay Grade	Rank
E4	Specialist or Corporal[a]
E5	Sergeant
E6	Staff Sergeant
E7	Sergeant First Class
E8	Master Sergeant
E9	Sergeant Major

[a]Most Soldiers at the E4 level are specialists; however, a small number are corporals. Specialists are not NCOs; corporals are considered junior NCOs.

In Phase III, the NCO21 project team identified measurement methods that could be used to assess the broadest range of the most critical KSAs across the two future eras. The team also identified measurement methods that could be used to assess NCO job performance. Knapp et al. (2002) documented the development and field testing of the predictor and criterion measures. In 2001, these instruments were used in a criterion-related validation data collection. The primary purpose of the validation effort was to determine which combination of KSA measures (i.e., performance predictors) best predicts important aspects of NCO performance (i.e., performance criteria).

NCO21 Predictor Measures

The NCO21 KSAs identified in Phase II are listed and defined in Table 1.2.[1] Note that the KSA list includes entries that may also be viewed as performance requirements. This is because performance requirements at one pay grade (e.g., E5) become relevant KSAs for promotion to the next higher grade (e.g., E6).

The Phase II SMEs provided judgments regarding the relative importance of the KSAs for current and future time periods. Although all KSAs in the list can be viewed as relevant, these judgments were used to help determine the KSAs that were most critical to measure in the NCO21 validation research effort.

[1] Following Phase II, additional work was done on these KSAs to clarify each and distinguish among them. Thus, this listing differs slightly from that provided in Ford et al. (2000).

Table 1.2. NCO21 Knowledges, Skills, and Aptitudes (KSAs) and Performance Requirements

Items 1-11 can be viewed as KSAs (i.e., predictors) only.

1. *Conscientiousness/Dependability.* The general tendency to be trustworthy, reliable, planful, and accountable. A general willingness to accept responsibility.

2. *General Cognitive Aptitude.* Has the overall capacity to understand and interpret information that is being presented, the ability to identify problems and reason abstractly, and the capability to learn new things quickly and efficiently.

3. *Need for Achievement.* Is generally predisposed to have confidence in own abilities and to seek and enjoy positions of leadership and influence. Would typically demonstrate enthusiasm and energy, and strive for accomplishment and recognition in almost any situation.

4. *Emotional Stability.* Has the tendency to act rationally and to display a generally calm, even mood. Typically maintains composure and is not overly distraught by stressful situations.

5. *Working Memory.* Has the ability to maintain information in memory for short periods of time and to retrieve it accurately.

6. *Spatial Relations Aptitude.* Has the ability to mentally visualize the relative positions of objects in two-dimensional or three-dimensional space, and how they will be positioned if they are moved or rotated in different ways.

7. *Perceptual Speed and Accuracy.* Has the ability to recognize and interpret visual information quickly and accurately, particularly with regard to comparing similarities and differences among words, numbers, objects, or patterns, when presented simultaneously or one after the other.

8. *Psychomotor Aptitude.* Has the ability to coordinate the simultaneous movements of one's limbs (arms, legs), to operate single controls or to operate multiple controls simultaneously, and to make precise control adjustments that involve eye-hand coordination.

9. *Basic Math Facility.* Knows and applies addition, subtraction, multiplication, division, and simple mathematical formulas.

10. *Basic Electronics Knowledge.* Knows general information regarding electronic principles and electronics equipment operation and repair. Knows general facts and principles relevant for a wide variety of electronics related tasks, but does not necessarily have highly specific electronics knowledge required for a particular job.

11. *Basic Mechanical Knowledge.* Knows general information regarding mechanical principles, tools, and mechanical equipment operation and repair. Knows general facts and principles relevant for a wide variety of tasks that require technical knowledge, but does not necessarily have highly specific mechanical knowledge required for a particular job.

Table 1.2. NCO21 Knowledges, Skills, and Aptitudes (KSAs) and Performance Requirements (Continued)

The remaining items can be viewed as either KSAs (predictors) or performance requirements (criteria).

12. *Problem-Solving/Decision Making Skill.* Reacts to new problem situations by applying previous experience and previous education/training appropriately and effectively. Does not apply rules or strategies blindly. Assesses costs and benefits of alternative solutions and makes timely decisions even with incomplete information.

13. *Writing Skill.* Communicates thoughts, ideas, and information successfully to others through writing. Uses proper sentence structure including grammar, spelling, capitalization, and punctuation.

14. *Oral Communication Skill.* Speaks in a clear, organized, and logical manner. Communicates detailed information, instructions, or questions in an efficient and understandable way. Note that this skill refers to how well the individual can speak and communicate, not whether technical expertise is high or low.

15. *MOS/Occupation-Specific Knowledge and Skill.* Possesses the necessary technical knowledge and skill to perform Military Occupational Specialty (MOS)/occupation-specific technical tasks at the appropriate skill level. Stays informed of the latest developments in field.

16. *Common Task Knowledge and Skill.* Possesses the necessary knowledge and skill to perform common tasks at the appropriate skill level (e.g., land navigation, field survival techniques, and nuclear, biological, and chemical [NBC] protection).

17. *Safety Consciousness.* Follows safety guidelines and instructions. Checks the behavior of others to ensure compliance.

18. *Computer Skills.* Understands computer systems, operating systems (e.g., Unix, Windows NT, and Army specific systems) and applications. Can perform routine troubleshooting of computer systems and applications.

19. *Motivating, Leading, and Supporting Individual Subordinates.* Recognizes, encourages, and rewards effective performance of individual subordinates. Corrects unacceptable conduct. Communicates reasons for actions and listens effectively to subordinates one-on-one. Fosters loyalty and commitment.

20. *Directing, Monitoring, and Supervising Individual Subordinates.* Works with subordinates one-on-one to assign tasks and set individual goals for work and assignments. Ensures that assignments are clearly understood. Monitors individual subordinate performance and gives appropriate feedback.

21. *Training Others.* Evaluates and identifies individual or unit training needs. Institutes formal or informal programs to address training needs. Develops others by providing appropriate work experiences. Guides and tutors subordinates on technical matters.

22. *Relating to and Supporting Peers.* Treats peers in a courteous, respectful, and tactful manner. Provides help and assistance to others. Backs up and fills in for others when needed. Works effectively as a team member.

23. *Team Leadership.* Communicates team goals and organizes and rewards effective teamwork. Leads the team to adapt quickly when missions change and keeps team focused on new goals. Resolves conflicts among team members. Shares relevant information with team members.

24. *Concern for Soldier Quality of Life.* Is aware of subordinates' off-duty needs and constraints. Is sensitive to others' priorities, interests, and values, and tries to assist subordinates in making their personal and family life better.

25. *Cultural Tolerance.* Demonstrates tolerance and understanding of individuals from other cultural and social backgrounds, both in the context of the diversity of U.S. Army personnel and interactions with foreign nationals during deployments or when training for deployment.

Table 1.2. NCO21 Knowledges, Skills, and Aptitudes (KSAs) and Performance Requirements (Continued)

26. *Modeling Effective Performance.* Acts in ways that consistently serve as a model for what effective performance should be like, be it technical performance, military bearing, commitment to the Army, support for the Army mission, or performance under stressful or adverse conditions. Can consistently set an example for others to follow.

27. *Level of Effort and Initiative on the Job.* Demonstrates high effort in completing work. Takes independent action when necessary. Seeks out and willingly accepts responsibility and additional challenging assignments. Persists in carrying out difficult assignments and responsibilities.

28. *Adherence to Regulations, Policies, and Procedures.* Adheres to policies and follows prescribed procedures in carrying out duties and assignments.

29. *Level of Integrity and Discipline on the Job.* Maintains high ethical standards. Does not succumb to peer pressure to commit prohibited, harmful, or questionable acts. Demonstrates trustworthiness and exercises effective self-control. Understands and accepts the basic values of the Army and acts accordingly.

30. Adaptability. Can modify behavior or plans as necessary to reach goals or to adapt to changing goals. Is able to maintain effectiveness when environments, tasks, responsibilities, or personnel change. Easily commits to learning new things when the technology, mission, or situation requires it.

31. *Physical Fitness.* Meets Army standards for weight, physical fitness, and strength. Maintains health and fitness to meet deployability and field requirements as well as the physical demands of the daily job.

32. *Military Presence.* Presents a positive and professional image of self and the Army even when off duty. Maintains proper military appearance.

*33. *Information Management.* Effectively monitors, interprets, and redistributes digital display information (as well as printed and orally delivered information) from multiple sources to multiple recipients. Sorts, classifies, combines, excludes, and presents information so that it is useable by others. Does not readily succumb to information overload.

*34. *Selfless Service Orientation.* Commits to the greater good of the team or group. Puts organizational goals ahead of individual goals as required.

*35. *General Self-Management Skill.* Uses appropriate strategies to self-manage the full range of own work and non-work responsibilities (e.g., work assignments, personal finances, family). Such strategies include setting both long- and short-term goals, allocation of effort and personal resources to goal priorities, and assessing one's own performance. Works effectively without direct supervision, but seeks help and advice from others when appropriate.

*36. *Self-Directed Learning Skill.* Has a clear goal of maintaining continuous learning and training over entire career. Is proficient at determining personal training needs, planning education and training experiences to meet them, and evaluating own training success. Uses efficient personal learning strategies (e.g., organizing the material to be learned, and practicing the new skills in an appropriate context).

*37. *Knowledge of the Inter-Relatedness of Units.* Is capable of analyzing how goals and operations of own unit are inter-related with other units and systems, and how one unit's actions affect the performance of other units. Can see the larger strategic picture and interpret how one's own unit relates to it.

*38. *Management and Coordination of Multiple Battlefield Functions.* Can individually apply and effectively integrate and coordinate multiple battlefield functions such as direct and indirect fires, communications, intelligence, and combat service support to achieve tactical goals.

Note. KSAs/performance requirements that are particularly relevant to one or both future eras, but not necessarily for the baseline era, are noted with an asterisk.

The project team identified eight predictor measures for use in the NCO21 project (see Table 1.3). The Armed Services Vocational Aptitude Battery (ASVAB) is a pre-enlistment test for which all Soldiers have archival scores. The Assessment of Individual Motivation (AIM) and the Biographical Information Questionnaire (BIQ) are operational tests used in the Army for other purposes. The BIQ is actually a compilation of multiple measures. Experimental versions of both the AIM and BIQ were prepared for use in the present research. The Situational Judgment Test (SJT), the SJT's close cousin (SJT-X), the Experience and Activities Record (ExAct), and a semi-structured interview were developed specifically for this project. Most of these instruments, however, made use of relevant, previously developed materials and items. Finally, the Personnel File Form (PFF21) was used to collect information that could be used to simulate current promotion system selection factors (e.g., awards and medals, civilian and military education).

Table 1.3. NCO21 Research Program Predictor and Criterion Measures

Predictors

- Personnel File Form-21 (PFF21) – archival information collected via self-report
- Situational Judgment Test (SJT)
- Situational Judgment Test-Experimental (SJT-X)
- Assessment of Individual Motivation (AIM)
- Biographical Information Questionnaire (BIQ)
- Semi-Structured Interview
- Experiences and Activities Record (ExAct)
- Armed Services Vocational Aptitude Battery (ASVAB) – archival

Criteria

- Observed Current Performance Rating Scales (supervisor ratings)
- Expected Future Performance Rating Scales (supervisor ratings)
- [Computerized simulation – data collected in another sample; to be reported separately]

Table 1.4 shows the predictor measures and indicates which of the 38 NCO21 KSAs are assessed by each. A checkmark indicates that the KSA is explicitly targeted by the instrument. An "X" indicates we would expect scores on the measure to correlate with direct measures of the KSA, even though the KSA is not explicitly targeted.

Only three KSAs have no coverage, either directly or indirectly. These are either low priority KSAs as identified by the Phase II expert panels (e.g., Safety Consciousness) or ones that would require very different measurement strategies than those that were adopted (e.g., Psychomotor Aptitude). A number of the higher priority KSAs are, however, addressed by multiple predictor measures.

NCO21 Criterion Measures

Phase II of the NCO21 project did not attempt to delineate specific task requirements for future NCOs, nor did it attempt to differentiate explicitly among performance requirements across NCO grades and time periods. Even with unlimited resources, it simply would not have been possible to abstract such specific predictions from the aggregate discussions and forecasts pertaining to the future Army. Phase II did, however, result in the identification of a set of forecasted future NCO performance requirements. Although still substantive in nature, these expected future requirements were defined more generally than specific task responsibilities, which cannot be forecasted with any degree of certainty. Descriptions of the sets of future performance requirements and the procedures by which they were generated are described in the Phase II report (Ford et al., 2000). Because performance at the E4 and E5 levels can be used to evaluate promotion potential, these performance requirements are included in the KSA set listed in Table 1.2 (see items 12-38).

Table 1.4. Measurement Methods by KSAs

KSA	Measurement Method						
	PFF21	SJT	AIM	BIQ	Interview	ExAct	ASVAB
General Cognitive Aptitude		X			X		
Working Memory							X
Basic Math Facility							
Basic Electronics Knowledge							
Basic Mechanical Knowledge							
Spatial Relations Aptitude							[a]
Perceptual Speed & Accuracy							[a]
Psychomotor Aptitude							
Problem-Solving/Decision Making							X
Information Management							X
Writing Skill	X						X
Oral Communication Skill							
MOS-Specific Knowledge & Skill	X						X
Common Task Knowledge & Skill	X						X
Safety Consciousness							
Computer Skills							
Knowledge of the Inter-Relatedness of Units		SJT-X					X
Management and Coordination of Multiple Battlefield Functions							X
Motivating, Leading, and Supporting Individual Subordinates			X		[b]		X
Directing, Monitoring, and Supervising Individual Subordinates			X				X
Training Others			X		[b]		X
Modeling Effective Performance	X	X	X			X	
Relating to and Supporting Peers							
Team Leadership					[b]		
Concern for Soldier Quality of Life							
Cultural Tolerance							

Table 1.4. Measurement Methods by KSAs (Continued)

KSA	Measurement Method						
	PFF21	SJT	AIM	BIQ	Interview	ExAct	ASVAB
Selfless Service Orientation							
Level of Effort and Initiative on the Job	X		X				
Need for Achievement							
Conscientiousness/Dependability							
Adherence to Regulations, Policies, and Procedures			X	X			
Level of Integrity and Discipline on the Job			X	X			
Emotional Stability							
Adaptability							
General Self-Management Skill							
Self-Directed Learning Skill	X						
Physical Fitness			X				
Military Presence							

Note. = designed to measure; X = expected to correlate.

[a]Spatial relations and perceptual speed and accuracy are measured by the Assembling Objects subtest which is now included as an experimental test on the CAT-ASVAB.

[b]Several KSAs were combined for measurement via the interview.

The primary criterion measures were two sets of instruments designed to collect performance information from supervisors. The Observed Performance Rating Scales cover all 27 NCO21 performance requirements. The 27 performance requirements, however, were consolidated into a more manageable set of 19 areas to be rated. The Expected Future Performance Rating Scales are not intended to measure the specific performance requirements, per se. Rather, they ask for evaluations of overall performance, given specific sets of alternative conditions expected to be characteristic of the future Army.

Under a separate contract effort, researchers from Aptima Human-Centered Engineering, Inc. developed a computer-based simulation that was also used as a criterion measure for some of the validation research participants. One goal of the developers was to assess at least two futuristic performance requirements that the supervisor ratings of current performance do not capture well (i.e., *Knowledge of the Inter-Relatedness of Units, Management/Coordination of Multiple Battlefield Functions)*. At the time of the criterion-related validation data collection effort, however, the simulation was in fairly early stages of development. Therefore, data were collected on only a small subset of the NCO21 validation research participants. Additional data collections that include the Aptima simulation, as well as most of the NCO21 predictor and criterion measures, were conducted in 2002. The Aptima simulation, data collections, and analysis will be described in a report prepared by Aptima (Hess et al., 2002).

Criterion-Related Validation

We used a concurrent design, collecting both predictor and criterion data from sergeants (grade E5) and staff sergeants (grade E6). To allow us to understand the distributional characteristics of the predictors in a key target sample (grade E4), the predictors were

administered to specialists/corporals as well. Table 1.5 summarizes some of the major research questions we addressed in the analysis of these data.

Table 1.5. Summary of Major Research Questions

• What is the psychometric quality of the predictor and criterion measures?
• What are the relations among the measures within each domain?
• What are the major dimensions of performance?
• To what extent does performance on the predictors relate to performance on various aspects of the job?
• What combination of predictors best predicts job performance?
• How does the best combination of predictors compare to the current set of predictors?

Note that the data collection design is limited in several ways. First, the concurrent design complicates our understanding of how predictors that are likely influenced by experience (e.g., the ExAct and the SJT) will work in a longitudinal situation. Second, we are interested in predicting performance in the future Army but are using Soldiers in the present Army in our research. Thus, we have to be concerned about how well we have understood and captured future conditions and requirements. At least one other limitation has to do with the fact that data were collected in a for-research-only environment. Threats to measurement accuracy that one could expect in an operational environment (e.g., "faking good" on the temperament measures) were likely not present.

Recommendations

Although there is some limited discussion of implementation-related issues, this is not the focus of the present report. Ideas and specific recommendations for implementation are discussed in a companion report (Knapp et al., 2003). Those recommendations will be based on results of the validation research, reactions to the instruments by Soldiers in the field, and input from Army stakeholders. We hope the suggestions will help address the complicated myriad of factors related to making a change to the Army's promotion processes (e.g., resource constraints, high volume of personnel actions).

Overview of Report

With Chapter 1 as background, subsequent chapters of this report focus on details of the NCO21 concurrent criterion-related validation effort. Chapter 2 presents administrative details of the data collection. Chapter 3 describes the psychometric characteristics of the ratings criterion measures. Chapters 4 through 8 discuss the scores, psychometric characteristics, zero-order validity estimates, and differential prediction analyses associated with each of the predictor instruments. Chapter 9 presents cross-instrument analyses that include the relationships among the predictors and criterion-related validity estimates. It includes a discussion of the findings as well as a more detailed discussion of caveats associated with the research design. Finally, Chapter 10 summarizes the technical findings of the NCO21 research program.

CHAPTER 2: VALIDATION DATA COLLECTION AND DATABASE DEVELOPMENT

Deirdre J. Knapp, Ani S. DiFazio, Laura A. Ford, and Dan J. Putka
HumRRO

This chapter describes the NCO21 criterion-related validation data collection, development of the analysis database, and final sample sizes following data cleaning and imputation.

Data Collection

Validation data were collected from April through August, 2001 at seven Army installations.

- Fort Bragg, NC
- Fort Campbell, KY
- Fort Carson, CO
- Fort Hood, TX
- Fort Lewis, WA
- Fort Riley, KS
- Fort Stewart, GA

The goal was to collect complete predictor data for E4 Soldiers, complete predictor and criterion data for E5 Soldiers, and partial predictor data (all except the interview) and complete criterion data for E6 Soldiers.

Data Collection Sites

Through ARI's formal troop support process, we requested a total of 2,455 Soldiers—along with two supervisors for each of the E5 and E6 Soldiers—to participate in the data collection. Actual troop support averaged about 77% of the requested numbers (n = 1,893 E4-E6 Soldiers). Performance ratings were collected from 1,022 supervisors.

Overview of On-Site Data Collection Activities

E4, E5, and E6 participants were scheduled for a 3-hour paper-and-pencil test session. Supervisors of the E5 and E6 Soldiers were asked to report to a separate location to provide performance ratings. E4 and E5 Soldiers were given the semi-structured interview in one of two ways. In some cases, Soldiers were scheduled for individual 45-minute sessions. Alternatively, Soldiers were taken from their paper-and-pencil test session to complete the interview, and then returned to their test session to finish testing.

A small sample of Soldiers (n = 24) at Fort Stewart completed the computerized simulation criterion measure developed by Aptima Human-Centered Engineering. These Soldiers also participated in the NCO21 data collection during the same time period. Aptima researchers collected additional simulation data, along with a subset of the NCO21 measures, from two sites in the spring of 2002. As mentioned in Chapter 1, the Aptima simulation research sample and associated analysis results will be the subject of a separate report (Hess et al., 2002).

The E4/E5/E6 Soldier and supervisor sessions involved the same initial steps. The data collection team introduced themselves, gave a brief project briefing, read a Privacy Act statement, and asked participants to complete a short Background Information Form. The Background Information Form asked for basic identifying information such as name, social security number (SSN), pay grade, and project identification code.

A list of instruments given in the Soldier paper-and-pencil sessions is provided in Table 2.1. For the most part, the E4-E6 Soldiers got the same forms in the 3-hour test session. The exception is that only the E6 participants took the SJT-X.

Table 2.1. Instruments Administered in Soldier Paper-and-Pencil Test Sessions

- Background Information Form
- Experiences and Activities Record (ExAct)
- Personnel File Form-21 (PFF21)
- Situational Judgment Test (SJT)
- SJT-X *(E6 Soldiers only)*
- Assessment of Individual Motivation (AIM)
- Biographical Information Questionnaire (BIQ)

Staff Training

HumRRO and ARI personnel served as test administrators. A data collection manual was developed that included information about how to prepare for and conduct the various data collection activities. This manual included sections containing the following information.

- Test schedules (e.g., timing and ordering of administration)
- Test and data security procedures
- Instructions for preparing the Soldier and supervisor "packets" that contained the forms to be completed by research participants
- Instructions for in-processing participants (e.g., assigning identification numbers, giving a project briefing and reading the Privacy Act statement)
- Instructions for administering the paper-and-pencil instruments
- Information about the Soldier interviews
- Instructions for identifying, in-processing, and training supervisor raters
- Data documentation and control procedures (e.g., instructions for maintaining rosters and logs and conducting on-site data quality checks on the various instruments)

In addition to reviewing the manual, data collection staff also participated in a half-day training program prior to collecting project data. This training reviewed and supplemented the material provided in the written manual.

Staff members serving as Test Site Managers or Interview Managers participated in another half-day of training that focused on their additional responsibilities. Test Site Managers were responsible for overall supervision and management of all data collection activities at their site. With the assistance of at least one other person, the Interview Manager at each site was responsible for training and overseeing the interviewers. A separate interviewer training guide manual was developed for the Interview Managers.

Interview Training and Administration

In addition to the E4-E6 Soldiers and their supervisors, participating Army installations were asked to provide 10 senior NCOs to participate as interviewers. At the beginning of each data collection, the NCO interviewers participated in a half-day training session. Interviewer training involved the following elements.

- NCO21 project briefing
- Discussion of the benefits of a structured interview
- Review of the interview components (performance areas, question bank, performance area rating scales)
- Discussion of the interview process (selecting and preparing questions, conducting the interview, evaluating the interviewee)
- Practice administering the interview and evaluating Soldiers

The NCOs were assigned to two-person interview teams, allowing up to five Soldiers to be interviewed at any given time. The senior NCO in each pair was designated the lead interviewer and the other NCO was the recorder. These roles had specific responsibilities (e.g., the lead interviewer had the final say in which questions would be asked, the recorder summarized and calculated final ratings). At the end of the data collection period, the NCO interviewers were asked to complete an evaluation form to collect information on their reactions to and ideas about the structured interview.

Supervisor Rating Sessions

In addition to the project briefing and Privacy Act statement, in-processing of supervisors included completing a rating card. Each card was used to list the names and identification codes of up to five Soldiers the supervisor would rate. Supervisors who could rate more than five Soldiers participating in the data collection were given a second card to complete.

Rater training involved (a) familiarizing the supervisors with the contents of the rating scales, (b) demonstrating how to use the rating cards when more than one Soldier was being rated, and (c) cautioning supervisors about common rating errors (e.g., halo, leniency, central tendency). Instructions were provided both orally and in writing. As supervisors completed their ratings, the ratings administrator provided additional coaching as needed (e.g., reminding supervisors to read the full definition for each performance area, pointing out ratings that seemed to reflect rating errors [such as uniformly high ratings across rating areas and ratees]).

A major responsibility of the ratings administrator was to ensure that two supervisors rated each E5 and E6 Soldier participating in the research. This proved quite difficult, because the supervisors did not generally report with their Soldiers as requested. Accordingly, during their in-processing, Soldiers were asked to identify and provide contact information for two supervisors. The ratings administrator then worked with installation tasking personnel to locate missing supervisors and schedule them for a rating session. After the first couple of data collections, we determined that there was a serious possibility of having insufficient criterion data to support the needs of the research. We therefore developed a procedure for collecting ratings from supervisors who could not participate in a face-to-face rating session while our staff was on-site.

Specifically, a "mail-back" procedure was devised to maximize the number of supervisor raters participating in the data collection. The mail-back supervisor packets included (a) a cover letter signed (when possible) by a senior officer from the installation, (b) a description of the project, (c) rating forms, (d) completed rating card, (e) supplemental instructions (in lieu of face-to-face training) for completing the materials, and (f) a metered return envelope addressed to HumRRO. The mail-back packets were distributed by installation personnel to those supervisors who were not able to participate in the research while data collection personnel were on site. Supervisors were generally given 2 weeks to complete and return the rating forms.

Database Construction

Initial Scanning and Scrubbing

The initial data processing and cleaning activities yielded five datasets that were provided to analysts for further data cleaning and imputation.

- All Soldier-level data (n = 1,881 to 1,892, depending on instrument; n = 525 for the SJT-X, which was administered only to E6 Soldiers)
- Soldier/supervisor-level performance ratings data
- Supervisor background information data (n = 1,022 raters)
- Soldier-level interview data (n = 946)
- Interviewer background information and evaluation feedback data (n = 58)

Addition of Archival Data

Soldier data on demographic (e.g., gender, race) and other variables (primarily scores on the ASVAB) were retrieved from the Army's automated Enlisted Master File (EMF) and added to the database. This was accomplished by matching SSNs from Soldiers in the NCO21 database to SSNs in the EMF.

Data Cleaning and Imputation

Several steps were taken to ensure the quality of the data gathered for the NCO21 validation effort. First, efforts were made to eliminate Soldiers' data on an instrument if more

than a given percentage of their responses was missing. For example, Soldiers who failed to respond to at least 90% of the items on the ExAct had their ExAct data dropped from further analyses. Due to variation in how instruments were structured, the approach varied slightly by instrument. Next, all self-report data were carefully screened for patterned or illogical response patterns. For example, responses to the AIM, BIQ, SJT, and ExAct were screened for Soldiers who repeatedly gave the same response to each item. Lastly, the problem logs that were kept during the collection of data for each instrument were reviewed to identify any Soldiers who might have provided questionable data. Details on these data-cleaning efforts, as well as the number of Soldiers who failed to meet these criteria, are provided in later chapters.

Given our goal of maintaining large sample sizes for purposes of validation, missing responses on instruments were imputed where possible.[2] One imputation method used was a multiple-regression based strategy, in which responses to an item on a given instrument were regressed on responses to all other items on that instrument. Missing responses were replaced with the predicted value from this regression equation plus random error (to avoid simply capitalizing on chance). The error that was added was drawn randomly from a normal distribution with a variance equal to the regression equation's squared standard error of estimate. This regression-based imputation method was used to impute (a) observed performance ratings, (b) SJT item scores, (c) ExAct item responses, and (d) some PFF21 item responses. Hot-deck imputation was used to impute other PFF21 responses. In this context, hot-deck imputation involved imputing Soldiers' missing responses based on the responses of Soldiers with similar characteristics (e.g., Soldiers of the same Career Management Field [CMF], pay grade, and gender). For continuously scaled responses, the mean response of similar Soldiers served as the estimate of the missing response. For categorically scaled responses, the response with the highest base rate among similar Soldiers served as the estimate of the missing response.

The amount of data requiring imputation was limited because of the data-cleaning steps aimed at eliminating Soldiers with many missing data. For example, less than 1% of ExAct data were imputed. Details on the imputation of missing responses, as well as the amount of data actually imputed, are provided in subsequent chapters.

Final Sample Sizes

Table 2.2 shows sample sizes following all data cleaning and imputation procedures, for the total sample and the key subgroups used in the analyses (pay grade, gender, race, and CMF category). Actual sample sizes, of course, vary by instrument and analysis.

[2] We did not impute missing data for the Expected Future Performance Rating Scales or interview scores, due in part to the small number of responses that constituted these instruments.

Table 2.2. Final Validation Sample Sizes by Subgroup (n = 1,889)

Subgroup	Pay Grade			
	E4	E5	E6	Total
Gender				
Male	365	770	498	1,633
Female	78	111	58	247
Race				
White	300	523	298	1,121
Black	92	246	184	522
Other	48	110	74	242
CMF Category				
Administration	70	85	60	215
Intelligence	21	37	21	79
Combat Operations	176	332	210	718
Logistics	143	290	170	603
Civil and Public Affairs	12	80	65	127
Communications	24	59	31	114
Total	449	885	557	

Summary

Data were collected from roughly 1,900 Soldiers in grades E4, E5, and E6 and from their supervisors at seven Army installations. Every effort was made to collect and prepare the NCO21 predictor and criterion data in a manner that would yield an accurate database with maximum sample sizes. Data collectors, NCO interviewers, and supervisor raters were carefully trained. Data collection staff monitored Soldiers, supervisors, and the NCO interviewers on-site to correct problems inasmuch as possible as they occurred. A process for collecting supervisor ratings through a mail-back procedure was successfully used to maximize the percentage of Soldiers for whom we collected criterion data. Once returned to HumRRO, data from the various predictor instruments, supervisor ratings, and the EMF were meticulously matched and merged. Numerous quality checks helped to ensure accuracy and imputation procedures were judiciously applied to maximize sample sizes.

The derivation of scores on the various instruments, which are included in the final database, is described in the following chapters. The final database, including all item-level and composite scores, has been documented and archived.

CHAPTER 3: SUPERVISOR RATINGS

Christopher E. Sager, Dan J. Putka, and Rodney A. McCloy
HumRRO

Overview

Two rating instruments were developed as criterion measures—one to assess observed job performance and another to forecast Soldier performance under expected future conditions. Specifically, the Observed Performance Rating Scales were used to collect supervisor ratings of subordinate Soldiers' typical behavior in areas covering a substantial portion of the job performance domain. These areas address all 27 NCO21 performance requirements listed in Table 1.2. The Expected Future Performance Rating Scales were used to obtain supervisor ratings of how well Soldiers could be expected to perform in scenarios describing conditions forecasted to occur in the future Army. These measures are based on the Project A model that conceptualizes job performance as a multidimensional construct comprising several distinct components (J. Campbell, McHenry, & Wise, 1990). The goal of these instruments is to describe and evaluate E5 and E6 Soldiers on requirements that constitute effective performance common to all Army jobs. Previous research has referred to such performance requirements as "Army-wide" criterion factors (Borman, Motowidlo, Rose, & Hanser, 1985).

Instrument Description

Observed Performance Rating Scales

The Observed Performance Rating Scales, which are modeled after and derived largely from previous Army NCO research, were developed in three stages.[3] First, a rating scale was developed for each of the 27 performance requirements, overall effectiveness, and senior NCO potential. Next, the prototype instrument, accompanied by written rater instructions and oral training, was pilot tested (second stage) on three occasions and then field tested (third stage) at three Army posts.

Following the pilot test, we reduced the number of scales from 27 to 19 to make the rating task more reasonable. The reduction from 27 to 19 requirement-specific scales was based on (a) an *a priori* model developed during Phase II of this project, (b) exploratory factor analyses of ratings collected during the field test, and (c) discussions among HumRRO and ARI project staff. The result was the consolidation of 13 of the original 27 scales into 5 combined scales.

Each of the 19 requirement-specific scales consists of a (a) title of the performance requirement being rated; (b) one sentence description of the performance requirement; and (c) 7-point rating scale, with three sets of requirement-specific behavioral anchors for points 1 – 2, 3 – 5, and 6 – 7, respectively (see Appendix A). Participating E5 and E6 Soldiers were also rated by their supervisors on 7-point scales assessing overall performance and senior NCO (i.e., E7-E9)

[3] For a detailed description of the development of the Observed Performance Rating Scales and the Expected Performance Rating Scales, see *Development of Predictor and Criterion Measures for the NCO21 Research Program* (Knapp et al., 2002).

potential. The overall effectiveness scale includes three sets of behavioral anchors, and the senior potential scale shows anchors asking the extent to which the Soldier would be a bottom-level, adequate, or top-level performer as a senior NCO.

Expected Future Performance Rating Scales

The Expected Future Performance Rating Scales also were developed in three stages. There was a concern that if scales designed to assess expected future performance took on a format too similar to that of the Observed Performance Rating Scales, method variance would result in artificially high correlations between scales assessing observed and expected future performance. To minimize this problem, we used themes identified in the future-oriented job analysis (Ford et al., 2002) to develop six scenarios describing conditions NCOs would likely face in the future Army. Each scenario is between one third and one half of a page long and is followed by a 7-point scale on which the supervisor rates the subordinate's expected performance effectiveness in the predicted future condition. Similar to the Observed Performance Rating Scales, the Expected Future Performance Rating Scales were pilot tested and then administered as part of the field test. These scales appear in Appendix B.

Results

Sample Sizes

One goal for the criterion rating scales was to obtain ratings from two supervisors for each E5 and E6 Soldier who participated in a written group administration session. The validation data collection involved administering criterion and predictor measures to Soldiers and supervisors at seven locations. At each site we administered a face-to-face rater-training program and monitored the supervisor raters as they completed the rating instruments.[4] As discussed in Chapter 2, obtaining the desired two raters per Soldier proved very difficult, so we developed a mail-back version of the rating packages to maximize our sample size. Approximately 33% of the ratings were mailed back.

We conducted interrater reliability analyses that included and excluded the mail-back responses to determine whether they had a deleterious effect on the reliability of the composite ratings.[5] The intraclass correlation (ICC) reliability estimates for a single rater, ICC(C,1), (McGraw & Wong, 1996) are provided in Table 3.1. The table shows there was little evidence that these ratings were any less reliable than those ratings collected on-site. Indeed, the mail-back ratings increased the instrument reliability estimates for three of the four instrument/grade combinations. Therefore, the mail-back ratings were included in all subsequent analyses.

[4] Research staff members trained each wave of supervisors. Thus, rater training occurred multiple times at each site.
[5] Development of the composite scores is discussed in later sections.

Table 3.1. Reliability Estimates for the Observed Performance Composite and Expected Future Performance Composite when Excluding and Including Mail-Back Responses

	Observed Performance		Expected Future Performance	
	E5	E6	E5	E6
Without Mail-Backs	.44	.47	.30	.39
With Mail-Backs	.45	.50	.31	.37

Table 3.2 presents sample sizes for the supervisor ratings following data preparation (i.e., matching Soldier predictor data to criterion data, cleaning, and imputation).[6] This table shows that 608 E5 and 393 E6 Soldiers have ratings of observed performance from at least one supervisor, which represents 68.7% of the E5 and 70.6% of the E6 Soldiers in the final validation sample. Similarly, 69.3% of the E5 and 71.6% of the E6 Soldiers in the sample have expected future performance ratings from at least one supervisor. Table 3.2 also shows the number of Soldiers rated by one or more supervisors. For example, 315 E5 Soldiers have ratings of observed performance from only one supervisor, and 261 E5 Soldiers have ratings of observed performance from two supervisors. Finally, the table shows the range of predictor/criterion matches. For example, for one predictor, only 471 of the 608 E5 Soldiers with criterion scores have predictor scores; for another predictor, all 608 E5 Soldiers have predictor scores.

Table 3.2. Final Sample Sizes for Supervisor Ratings by Pay Grade

Criterion	E5 Soldiers		E6 Soldiers	
	Observed Performance	Expected Future Performance	Observed Performance	Expected Future Performance
Number of Soldiers with supervisor ratings	608	613	393	399
Number of supervisor ratings per Soldier				
1	315	313	198	210
2	261	271	175	166
3	30	27	17	20
4+	2	2	3	3
Number of predictor-supervisor rating matches	471-608	474-613	341-393	346-399

Observed Performance Rating Scales

Data Preparation

Preparation of the observed performance ratings involved four steps: (a) eliminating from further analysis scales for which the response rate was too low, (b) eliminating supervisor/Soldier pairs in which the supervisor had worked with the Soldier for less than 1

[6] General data preparation is discussed in Chapter 2; data preparation specific to the ratings is discussed in the following sections of this chapter.

month, (c) eliminating pairs in which the supervisor rated the Soldier on too few scales, and (d) imputing missing values for the remaining supervisor/Soldier pairs. For purposes of data preparation, a rating was declared missing if no response options were marked or if the "Cannot Rate" option was selected.

Out of 21 scales (i.e., 19 performance-requirement scales, 1 overall effectiveness scale, and 1 senior NCO potential scale), only Scale 17 (*Coordinating Multiple Units and Battlefield Functions*) was eliminated from further analysis because of a low response rate. For this scale, 381 (22.8%) of 1,668 supervisor/Soldier pairs had missing values. This is not an unexpected result, given that Scale 17 covers an area that NCOs are predicted to perform more frequently in the future. Currently, however, E5 and E6 Soldiers have few opportunities to demonstrate performance in this domain. The number of missing values for the remaining scales ranged from 23 (1.4%) to 220 (13.2%).

Another 120 (i.e., 7.2%) supervisor/Soldier pairs were dropped from further analysis because the supervisor either (a) had not worked with the Soldier for at least 1 month or (b) did not rate the Soldier on at least 90% of the remaining items (i.e., 18 out of 20). Because many Soldiers were rated by more than one supervisor, however, the loss of 120 pairs resulted in only a 3.2% drop in the number of Soldiers with at least one set of observed performance ratings (i.e., from $n = 1{,}035$ to $n = 1{,}001$).

Finally, the regression-based approach to imputation described in Chapter 2 was used to impute missing values for the remaining 1,548 supervisor/Soldier pairs. Specifically, for a given supervisor/Soldier pair, we used the ratings the supervisor did provide to predict the missing ratings. The 1,548 supervisor/subordinate pairs involve 30,960 scale-level ratings, only 712 (2.3%) of which required imputation.

Descriptive Statistics

As shown in Table 3.2, these analyses include a total of 1,001 Soldiers ($n_{E5} = 608$; $n_{E6} = 393$) who were each rated by at least one supervisor. Table 3.3 shows the descriptive statistics for the E5 and E6 Observed Performance Rating Scale scores. For each Soldier, each rating scale score is based on all supervisors who rated that Soldier. For example, if the Soldier was rated by one supervisor, the Soldier's score on Scale 3 (Computer Skills) is the rating made by that single supervisor; if the Soldier was rated by two or more supervisors, the Soldier's score on Scale 3 is the mean rating of those two or more supervisors. The mean rating scores in this table suggest some leniency in the ratings. The standard deviations, however, indicate that supervisors were able to discriminate among Soldiers on each scale. The last row of this table shows the descriptive statistics for an observed performance composite score. For each Soldier, the composite score is based on the mean across the 18 requirement-specific scale ratings.

As will be the case for all the instruments described in this report, the following sections present descriptive statistics for the total sample and for subgroups based on pay grade, race, gender, and CMF. Effect sizes that show the magnitude and statistical significance of subgroup differences in mean scores are also reported. Subgroup difference are of general interest for all the instruments, but particularly for the experimental predictors that will be discussed in subsequent chapters. Race and gender difference are of particular concern because selection and

promotion systems should minimize adverse impact against racial minority groups and women. Chapters describing predictors will also examine differential prediction (slope difference) across subgroups. This is a standard second step (after examining mean score [intercept] differences on the criterion) for evaluating test bias (Cleary, 1968). Pay grade and CMF differences are of interest in part because performance differences might suggest differences in how the measures might be best utilized in a promotion system. That is, even though the Army's current promotion system is the same promotions to E5 and E6, regardless of job type, improvements might be gained by tailoring the system to each pay grade and/or across job types.

Raw and conditional means. Descriptive statistics for the observed performance rating composite are reported by subgroup (pay grade, race, gender, and CMF cluster) in Tables 3.4 and 3.5. Table 3.4 reports sample sizes, means, standard deviations, and effect sizes by pay grade, as well as by gender and race (within each pay grade). Table 3.5 reports sample sizes, means, standard deviations, and effect sizes by CMF cluster (within each pay grade). Raw and conditional statistics are reported in all tables. Effect sizes are reported only for comparisons in which each subgroup contained at least 20 individuals.

Conditional means and effect sizes offer the benefit of reflecting estimated differences between subgroups while holding other grouping variables constant. For example, comparing the conditional means of gender removes differences between males and females that are due to differences in composition of the two samples in terms of race, pay grade, and CMF cluster. See Appendix C for a discussion of conditional means, effect sizes, and their calculation.

Raw and conditional effect sizes. Raw effect sizes reported in Table 3.4 were calculated by taking the mean of the non-referent group (e.g., females, blacks) minus the mean of the referent group (e.g., males, whites), and dividing the resulting quantity by the standard deviation of the referent group. Raw effect sizes reported in Table 3.5 were calculated by taking the mean of the higher-numbered CMF cluster (e.g., 2. Intelligence) minus the mean of the lower-numbered CMF cluster (e.g., 1. Administration) and dividing the resulting quantity by the overall standard deviation in the pay grade of interest.

Conditional effect sizes were calculated by taking the conditional mean of the non-referent group minus the conditional mean of the referent group, and dividing the resulting quantity by the pooled standard deviation for the referent group (within each pay grade). Conditional effect sizes reported in the second table of each pair were calculated by taking the conditional mean of the higher numbered CMF cluster minus the conditional mean of the lower numbered CMF cluster, and dividing the resulting quantity by the *overall* pooled standard deviation (within each pay grade).

Given their greater experience and higher rank, it is not surprising that E6 Soldiers had significantly higher mean performance ratings than E5 Soldiers. It is also reassuring that there were no significant differences in performance ratings obtained for the demographic subgroups (gender and race). E5 Soldiers in the Administration CMF were consistently rated higher than E5 Soldiers in other CMF. There were significant differences between some other CMF. There was no obvious pattern nor was one expected.

Table 3.3. Descriptive Statistics and Interrater Reliability Estimates for Observed Performance Ratings

Scale/Composite		E5 Soldiers				E6 Soldiers			
		M	*SD*	ICC(C,1)	ICC(C,*k*)	*M*	*SD*	ICC(C,1)	ICC(C,*k*)
1	MOS/Occupation-Specific Knowledge and Skill	5.10	1.15	.42	.51	5.55	1.11	.42	.51
2	Common Task Knowledge and Skill	5.33	1.06	.27	.35	5.72	1.02	.31	.40
3	Computer Skills	4.60	1.39	.31	.39	4.91	1.30	.31	.40
4	Writing Skill	4.54	1.21	.27	.35	4.92	1.09	.17	.24
5	Oral Communication Skill	4.98	1.23	.35	.44	5.29	1.08	.35	.44
6	Level of Effort/Initiative on the Job	5.16	1.35	.37	.46	5.52	1.22	.38	.47
7	Adaptability	4.90	1.20	.25	.33	5.22	1.12	.40	.49
8	Self-Management and Self-Directed Learning Skill	4.89	1.27	.30	.38	5.38	1.08	.40	.49
9	Demonstrated Integrity, Discipline, and Adherence to Army Procedures	5.45	1.19	.29	.37	5.83	1.02	.17	.24
10	Acting as a Role Model	4.95	1.40	.38	.47	5.30	1.32	.40	.49
11	Relating to and Supporting Peers	5.17	1.15	.23	.30	5.39	1.04	.38	.47
12	Cultural Tolerance	5.79	0.94	.04	.06	5.96	0.82	.04	.06
13	Selfless Service Orientation	5.16	1.22	.28	.36	5.48	1.06	.35	.44
14	Leadership Skills	4.85	1.28	.35	.44	5.39	1.15	.38	.47
15	Concern for Soldier Quality of Life	5.29	1.11	.28	.36	5.59	0.95	.21	.28
16	Training Others	4.83	1.24	.28	.36	5.25	1.14	.34	.43
18	Problem-Solving/Decision Making Skill	4.81	1.20	.31	.39	5.27	1.14	.30	.39
19	Information Management	4.72	1.17	.27	.35	5.20	1.10	.20	.27
20	Overall Effectiveness Rating	4.94	1.13	.34	.43	5.31	1.06	.44	.53
21	Senior NCO Potential Rating	4.82	1.39	.37	.46	5.36	1.24	.40	.49
Observed Performance Composite		5.03	0.84	.45	.53	5.40	0.74	.50	.59

Note. $n_{E5} = 608$; $n_{E6} = 393$.

Table 3.4. Subgroup Differences by Pay Grade, Gender, and Race for the Observed Performance Rating Composite

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E5										
Gender										
Female	74	4.89	0.90	-0.19	.127	60	4.64	0.91	-0.49	.049
Male	532	5.05	0.83			464	5.04	0.82		
Race										
Black	153	5.00	0.81	-0.01	.935	153	4.81	0.81	-0.09	.591
White	372	5.00	0.85			371	4.88	0.84		
E6										
Gender										
Female	36	5.29	0.83	-0.16	.373	30	5.00	0.87	-0.50	.074
Male	356	5.41	0.73			311	5.37	0.74		
Race										
Black	131	5.34	0.83	-0.16	.179	131	5.09	0.83	-0.28	.185
White	212	5.45	0.71			210	5.28	0.70		
Grade										
E6	393	5.40	0.74	0.44	<.001	341	5.18	0.75	0.41	.002
E5	608	5.03	0.84			524	4.84	0.83		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed t-tests of differences between subgroup means. Effect sizes are reported only for comparisons in which each subgroup contained at least 20 individuals.

Latent Structure and Composite Scores

Confirmatory and exploratory factor analyses of the Observed Performance Rating Scales were conducted to (a) determine the latent structure underlying these ratings and (b) develop observed performance scores for use in criterion-related validity analyses.[7] These analyses did not strongly support the presence of multiple factors. Therefore, a single observed performance composite score was calculated for each Soldier as the mean rating of all ratings received (i.e., the mean rating across all scales and supervisors). Because Soldiers received ratings from different numbers of supervisors (some from only one, others from four or more), the observed performance composite score for a given Soldier will be based upon $18*n_s$ data points, where n_s is the number of supervisors who rated the Soldier in question. The correlations among the 18 observed performance scales, the effectiveness and NCO potential scales, and the overall composite score appear in Table 3.6.

[7] *Coordination of Multiple Units and Battlefield Functions* was excluded from these analyses.

Table 3.5. Differences between CMF Clusters for the Observed Performance Ratings Composite

CMFC	*N*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E5 Soldiers												
1. ADM	39	33	5.28	5.35	0.73	0.76	—	-0.52	-1.58 **	-0.56 **	-0.50	-0.55
2. INT	21	21	4.75	4.92	0.99	1.01	-0.64 **	—	-1.06	-0.03	0.02	-0.02
3. CBO	235	200	5.05	4.04	0.82	0.82	-0.28	0.36	—	1.02 *	1.08 *	1.03
4. LOG	208	177	4.95	4.89	0.85	0.86	-0.39 *	0.25	-0.11 *	—	0.06	0.01
5. CPA	64	59	5.16	4.94	0.77	0.76	-0.15	0.49	0.13 *	0.24	—	-0.05
6. COM	40	34	4.97	4.90	0.90	0.79	-0.37	0.27	-0.09	0.02	-0.22	—
Overall	608		5.03		0.84							
E6 Soldiers												
1. ADM	34	32	5.41	5.42	0.91	0.94	—	.	-1.38 *	0.12	-0.24	.
2. INT	12	9	5.28	5.41	0.53	0.59	.	—	.	.	.	.
3. CBO	157	137	5.30	4.38	0.75	0.76	-0.16	.	—	1.50 **	1.14	.
4. LOG	121	104	5.52	5.51	0.71	0.71	0.15	.	0.30 *	—	-0.37	.
5. CPA	46	40	5.47	5.24	0.75	0.79	0.08	.	0.23	-0.07	—	.
6. COM	23	19	5.32	5.14	0.54	0.57	-0.12	.	0.04	-0.27	-0.20	—
Overall	393		5.40		0.74							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 3.6. Observed Performance Rating Intercorrelations

Scale/Composite		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	19	20	21	22	23
1	MOS/Occupation-Specific Knowledge and Skill	.	.62	.34	.44	.48	.61	.48	.49	.39	.48	.41	.21	.47	.56	.29	.55	.53	.54	.71	.68	.70	.74
2	Common Task Knowledge and Skill	.61	.	.37	.42	.47	.49	.51	.53	.42	.47	.38	.26	.38	.57	.31	.50	.50	.47	.64	.60	.68	.72
3	Computer Skills	.24	.26	.	.48	.34	.22	.36	.26	.14	.08	.18	.13	.19	.23	.20	.29	.34	.49	.33	.25	.40	.48
4	Writing Skill	.40	.40	.49	.	.56	.36	.45	.45	.31	.26	.37	.22	.32	.35	.24	.37	.42	.58	.50	.44	.58	.63
5	Oral Communication Skill	.45	.49	.28	.51	.	.39	.51	.50	.39	.37	.45	.35	.38	.49	.36	.50	.59	.53	.54	.55	.67	.71
6	Level of Effort/Initiative on the Job	.59	.55	.17	.37	.46	.	.46	.56	.56	.51	.42	.17	.58	.57	.32	.54	.49	.41	.68	.67	.68	.72
7	Adaptability	.54	.48	.18	.39	.48	.58	.	.55	.37	.38	.41	.28	.41	.55	.31	.50	.59	.54	.60	.57	.67	.72
8	Self-Management and Self-Directed Learning Skill	.52	.51	.29	.51	.51	.61	.52	.	.50	.49	.43	.28	.46	.57	.35	.50	.52	.46	.64	.60	.70	.74
9	Demonstrated Integrity, Discipline, and Adherence to Army Procedures	.46	.44	.11	.32	.37	.62	.46	.54	.	.50	.50	.31	.55	.49	.33	.43	.40	.33	.56	.54	.60	.65
10	Acting as a Role Model	.46	.45	.13	.34	.43	.61	.45	.59	.62	.	.38	.15	.38	.57	.28	.45	.36	.30	.56	.61	.56	.62
11	Relating to and Supporting Peers	.46	.46	.23	.38	.42	.53	.47	.50	.55	.49	.	.44	.53	.52	.49	.50	.47	.39	.53	.47	.63	.67
12	Cultural Tolerance	.25	.28	.10	.20	.25	.25	.31	.28	.32	.25	.39	.	.33	.25	.40	.29	.35	.34	.27	.27	.40	.45
13	Selfless Service Orientation	.44	.43	.13	.27	.30	.61	.45	.51	.61	.50	.52	.32	.	.53	.42	.48	.47	.39	.56	.54	.63	.68
14	Leadership Skills	.63	.58	.20	.43	.53	.64	.58	.65	.57	.61	.57	.30	.58	.	.40	.59	.55	.46	.74	.74	.73	.77
15	Concern for Soldier Quality of Life	.43	.44	.23	.35	.44	.47	.46	.48	.48	.41	.53	.40	.49	.57	.	.41	.40	.34	.37	.36	.50	.55
16	Training Others	.58	.57	.19	.41	.50	.56	.57	.52	.50	.46	.46	.23	.43	.65	.49	.	.60	.49	.67	.65	.70	.75
18	Problem-Solving/Decision Making Skill	.60	.55	.24	.47	.52	.55	.61	.57	.49	.46	.51	.29	.48	.66	.48	.55	.	.53	.66	.63	.71	.75
19	Information Management	.48	.43	.42	.53	.46	.45	.47	.50	.39	.38	.42	.27	.45	.51	.45	.49	.58	.	.57	.52	.67	.71
20	Overall Effectiveness Rating	.71	.62	.22	.48	.57	.71	.60	.68	.62	.64	.58	.27	.55	.76	.52	.68	.69	.55	.	.84	.	.84
21	Senior NCO Potential Rating	.62	.53	.23	.47	.53	.67	.56	.64	.57	.65	.53	.28	.54	.73	.50	.61	.64	.50	.80	.	.	.81
22	Observed Performance Composite with scale deleted	.70	.68	.32	.57	.63	.74	.68	.74	.67	.65	.67	.39	.64	.80	.64	.70	.74	.65	.	.	.	.
23	Observed Performance Composite	.73	.71	.40	.63	.68	.78	.72	.78	.71	.70	.71	.44	.69	.83	.69	.74	.77	.70	.84	.79	.	.

Note. Scale 17 (Coordination of Multiple Units and Battlefield Functions) excluded. $n_{E5} = 608$; $n_{E6} = 393$. Correlations for E5 Soldiers appear below the diagonal. Correlations for E6 Soldiers appear above the diagonal. All correlations were statistically significant, $p < .05$ (one-tailed).

In addition to the single composite, we generated six other composites based on a rational grouping of the 18 observed performance scales. These six composites served to preserve the notion of a multidimensional performance space by "manually overriding" the method variance believed to be driving the factor analyses toward a single-factor solution. Table 3.7 presents these composites and their constituent scales. Table 3.8 provides descriptive statistics and interrater reliability estimates for the six rational composite scores for E5 and E6 Soldiers. As one would expect, the reliability estimates tend to be a bit higher than the estimates for the single scales but not quite as high as the estimate for the overall composite rating. *Leadership: Consideration* and *Information Management* were rated least reliably by E5 and E6 Soldiers alike, whereas *Leadership: Structure* was one of the most reliably rated composites.

Table 3.7. Mapping of Observed Performance Rating Scales onto Factor Composites

Factor/Scale
Technical Performance
MOS/Occupation-Specific Knowledge and Skill
Common Task Knowledge and Skill
Leadership: Structure
Oral Communication Skill
Adaptability
Leadership Skills
Training Others
Problem-Solving/Decision Making Skill
Effort-Integrity-Selfless Service
Level of Effort/Initiative on the Job
Demonstrated Integrity, Discipline, and Adherence to Army Procedures
Selfless Service Orientation
Leadership: Consideration
Relating to and Supporting Peers
Cultural Tolerance
Concern for Soldier Quality of Life
Information Management
Computer Skills
Writing Skill
Information Management
Individual Self-Management
Self-Management and Self-Directed Learning Skill
Acting as a Role Model

Interrater Reliability

In addition to basic descriptive statistics, Table 3.3 shows the ICC reliability estimates for the scores on the Observed Performance Rating Scales (McGraw & Wong, 1996). The first step in generating each estimate was to create a subsample of Soldiers who had ratings from two or more supervisors. For each of these Soldiers, two supervisors were randomly selected; they were then labeled as raters 1 and 2, respectively. This allowed for the calculation of an ICC(C,2) reliability estimate for rating scores based on ratings from two supervisors (i.e., an estimate of the *consistency* of the ratings provided by k=2 raters). Next, the Spearman Brown prophecy formula (Crocker & Algina, 1986) was used to calculate ICC reliability estimates for rating scores based on ratings from k=1 and k=3 or more supervisors (k ranged to a maximum of 4 supervisors for E5 Soldiers and 5 supervisors for E6 Soldiers). Finally, given that a particular Soldier's scale score could be based on ratings provided by one, two, or occasionally more supervisors, our estimates of criterion reliability for the whole sample are weighted averages of ICC(C,1), ICC(C,2), and ICC(C,k) values based on the proportion of the sample that was rated by one, two, or k raters. These weighted reliability estimates are the ICC(C,k) values shown in Table 3.3. The observed performance composite interrater reliability estimates for E5 and E6 Soldiers were .53 and .59, respectively. These values are consistent with those typically found with performance ratings (Viswesvaran, Ones, & Schmidt, 1996).

Table 3.8. Descriptive Statistics and Interrater Reliability Estimates for Observed Performance Factor Composites

Factor Composite	E5 Soldiers				E6 Soldiers			
	M	*SD*	ICC(C,1)	ICC(C,*k*)	*M*	*SD*	ICC(C,1)	ICC(C,*k*)
1 Technical Performance	5.22	0.99	.39	.48	5.63	0.96	.43	.52
2 Leadership: Structure	4.87	0.99	.41	.50	5.28	0.90	.49	.58
3 Effort/Integrity/Selfless Service	5.25	1.08	.42	.51	5.61	0.93	.40	.49
4 Leadership: Consideration	5.42	0.85	.27	.35	5.65	0.74	.31	.40
5 Information Management	4.62	1.01	.37	.45	5.01	0.96	.32	.41
6 Individual Self-Management	4.92	1.19	.41	.50	5.34	1.04	.45	.54

Note. n_{E5} = 608; n_{E6} = 393; *M* = the mean of the mean scores on the Observed Performance Rating Scales associated with each composite; *SD* = the standard deviation of the mean scores on the Observed Performance Rating Scales associated with each factor composite.

Future Performance Rating Scales

Data Preparation

Preparation of the expected future performance ratings involved two steps: (a) eliminating supervisor/Soldier pairs in which the supervisor had worked with the Soldier for less than 1 month and (b) eliminating pairs in which the supervisor rated the Soldier on too few scales. A rating was declared missing if no response options were marked (the "Cannot Rate" option was not available for the expected future performance ratings). There are only six expected future performance scales. Therefore, if the supervisor failed to rate the Soldier on even one of the six scales, the supervisor/Soldier pair was eliminated for having more than 10% missing data. These two steps together resulted in a 2.6% reduction in supervisor/Soldier pairs

(i.e., from 1,642 to 1,599); however, because many Soldiers were rated by more than one supervisor, this resulted in only a 1.8% reduction in Soldiers with at least one complete set of expected future performance ratings (i.e., from 1,031 to 1,012). No imputation was necessary because the remaining pairs had no missing data.

Descriptive Statistics

Table 3.9 shows the descriptive statistics for the expected future performance E5 and E6 Soldier rating scores. As with the Observed Performance Rating Scales, each Soldier's rating scale score is based on all supervisors who rated that Soldier. Also, as seen in the Observed Performance Rating Scales, (a) the mean rating scores suggest some leniency (although less so than with the observed performance scales), and (b) the standard deviations indicate that supervisors were able to discriminate among Soldiers on each scale. The last row of this table shows the descriptive statistics for an expected future performance composite score. The composite score is calculated the same way as the score on the observed performance composite—as the mean across all scenario scale ratings received by a Soldier. Thus, each Soldier's expected future performance composite score will be the mean of $6*n_s$ ratings, where again n_s is the number of supervisors rating that Soldier.

Table 3.9. Descriptive Statistics and Interrater Reliability Estimates for Expected Future Performance Ratings

Scenario/Composite	E5 Soldiers				E6 Soldiers			
	M	*SD*	ICC(C,1)	ICC(C,*k*)	*M*	*SD*	ICC(C,1)	ICC(C,*k*)
1 Increased Requirements for Self-Direction and Self-Management	4.82	1.21	.30	.38	5.25	1.11	.26	.34
2 Use of Computers, Computerized, Equipment, and Digitized Operations	4.90	1.18	.20	.27	5.23	1.10	.21	.28
3 Increased Scope of Technical Skill Requirements	4.93	1.08	.18	.25	5.15	1.10	.16	.22
4 Increased Requirements for Broader Leadership Skills at Lower Levels	4.81	1.21	.29	.37	5.14	1.17	.36	.45
5 Need to Manage Multiple Operational Functions and Deal with the Inter-relatedness of Units	4.68	1.13	.18	.25	5.06	1.11	.28	.36
6 Mental and Physical Adaptability and Stamina	5.06	1.24	.31	.39	5.14	1.25	.36	.45
Expected Future Performance Composite	4.86	0.96	.31	.39	5.16	0.93	.37	.46

Note. $n_{E5} = 613$; $n_{E6} = 399$.

Descriptive statistics for the expected future performance composite are reported by subgroup in Tables 3.10 and 3.11. Raw and conditional statistics are reported in all tables. Effect sizes were reported only for comparisons in which each subgroup contained at least 20 individuals. As with the observed performance ratings, E6 Soldiers were rated higher than E5 Soldiers. For the expected future performance ratings, however, men tended to be rated higher than women at both pay grades. There were several differences between ratings obtained by Soldiers in different CMF, but no notable pattern.

Table 3.10. Subgroup Differences by Pay Grade, Gender, and Race for the Expected Future Performance Rating Scales Composite

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E5										
Gender										
Female	76	4.54	0.97	-0.39	.002	61	4.18	0.91	-0.76	.002
Male	535	4.91	0.95			465	4.91	0.96		
Race										
Black	158	4.85	0.93	0.01	.919	157	4.52	0.89	-0.04	.780
White	370	4.84	0.98			369	4.57	0.97		
E6										
Gender										
Female	36	4.78	1.24	-0.47	.011	30	4.49	1.36	-0.73	.008
Male	362	5.20	0.89			316	5.14	0.89		
Race										
Black	132	5.10	1.09	-0.15	.224	132	4.72	1.05	-0.23	.267
White	216	5.23	0.86			214	4.91	0.85		
Grade										
E6	399	5.16	0.93	0.31	<.001	346	4.81	0.93	0.28	.043
E5	613	4.86	0.96			526	4.55	0.95		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed t-tests of differences between subgroup means.

Scale Intercorrelations

Table 3.12 provides the correlations among the six future performance scales for E5 and E6 Soldiers, respectively. The patterns of covariation are quite similar across both grades, as are the mean intercorrelations (.60 for E5 Soldiers, .61 for E6 Soldiers). The correlations generally range from .50-.70, although the correlations of *Use of Technology* with *Broader Leadership* and *Adaptability and Stamina* are lower (.48 and .37 for E5 Soldiers, .47 and .46 for E6 Soldiers, respectively). Correlations of the factors with the composite score range from .70 to .87 for both pay grades. These scales were not factor analyzed because there was no hypothesized underlying latent structure beyond a single factor represented by the composite score.

Interrater Reliability

In addition to basic descriptive statistics, Table 3.9 shows the ICC reliability estimates for the expected future performance scores. These estimates were calculated the same way as the estimates for the Observed Performance Rating Scales. The expected future performance composite ICC (C,*k*) interrater reliability estimates for E5 and E6 Soldiers were .39 and .46, respectively. Although lower than the estimates obtained for the Observed Performance Rating Scales, these estimates are still consistent with past research (Viswesvaran et al., 1996).

Table 3.11. Differences between CMF Clusters for the Expected Future Performance Rating Scales Composite

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E5 Soldiers												
1. ADM	41	34	5.07	5.03	0.82	0.78	—	-0.15	-1.33*	-0.39	-0.67*	-0.48
2. INT	21	21	4.70	4.89	1.20	1.21	-0.38	—	-1.18	-0.25	-0.53	-0.34
3. CBO	235	199	4.95	3.77	0.97	0.99	-0.12	0.26	—	0.93	0.65	0.84
4. LOG	208	177	4.79	4.65	0.92	0.91	-0.29	0.10	-0.16	—	-0.28	-0.09
5. CPA	67	62	4.83	4.39	0.96	0.90	-0.25	0.13	-0.12	0.04	—	0.19
6. COM	39	33	4.65	4.57	1.05	0.97	-0.44*	-0.06	-0.31	-0.15	-0.19	—
Overall	613		4.86		0.96							
E6 Soldiers												
1. ADM	33	31	5.19	5.21	1.22	1.27	—	.	-1.23*	-0.01	-0.65*	-0.80**
2. INT	12	9	5.24	5.36	1.03	0.98	.	—	.	.	.	.
3. CBO	158	138	5.14	4.06	0.88	0.88	-0.05	.	—	1.22*	0.59	0.43
4. LOG	123	106	5.31	5.19	0.85	0.85	0.13	.	0.18	—	-0.63*	-0.79**
5. CPA	48	41	5.02	4.60	0.98	1.00	-0.18	.	-0.13	-0.31*	—	-0.15
6. COM	25	21	4.73	4.46	1.01	1.02	-0.49	.	-0.44	-0.62 **	-0.31	—
Overall	399		5.16		0.93							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 3.12. Correlations among the Expected Future Performance Rating Scales and Future Performance Composite

Scale/Composite	EFP1	EFP2	EFP3	EFP4	EFP5	EFP6	EFP Composite w/o scale	EFP Composite
EFP: Scenario 1 Self-Direction	.	.57	.68	.68	.63	.63	.78	.85
EFP: Scenario 2 Use of Technology	.51	.	.61	.47	.48	.46	.61	.73
EFP: Scenario 3 Scope Technical Skills	.65	.53	.	.67	.72	.61	.80	.87
EFP: Scenario 4 Broader Leadership	.73	.48	.68	.	.69	.62	.76	.84
EFP: Scenario 5 Manage Multi Operational Functions	.67	.54	.68	.72	.	.58	.75	.83
EFP: Scenario 6 Adaptability and Stamina	.64	.37	.57	.61	.63	.	.70	.80
Expected Future Performance Ratings Composite with scale deleted	.79	.56	.76	.79	.80	.68	.	.
Expected Future Performance Ratings Composite	.86	.70	.83	.86	.86	.79	.	.

Note. $n_{E5} = 613$; $n_{E6} = 399$. Correlations for E5 Soldiers appear below the diagonal. Correlations for E6 Soldiers appear above the diagonal. Composites computed both with and without the applicable scenario scales. All correlations significant at $p < .001$.

Correlations of Observed Performance with Future Performance

Observed Factor Scores with Future Performance

One remaining question concerns the degree to which various dimensions of observed performance relate to expected future performance. Correlations between the six rational observed performance factor scores and the six expected future performance scales for E5 and E6 Soldiers appear in Table 3.13. This table also contains the observed and future composites. As with the future performance factors, the patterns of covariation are quite similar across pay grades, as are the mean intercorrelations (.54 for E5 Soldiers, .53 for E6 Soldiers). In addition, the covariation pattern is sensible, with high correlations where one might expect (e.g., *Information Management* with *Use of Technology*, *Individual Self-Management* with *Self-Direction*). The correlations exhibit a rather wide range of values, with the lowest correlations in the .30s (e.g., *Information Management* with *Adaptability and Stamina*; *Integrity-Selfless Service* and *Leadership: Consideration* with *Use of Technology*) and the highest in the .70s (e.g., *Leadership: Structure* with *Self-Direction*). The observed performance factor *Leadership: Structure* exhibited strong correlations with the future performance composite for both pay grades (.78 for E5 Soldiers, .77 for E6 Soldiers), whereas *Leadership: Consideration* did not correlate as highly (.58 for E5 Soldiers, .49 for E6 Soldiers). For both pay grades, the two performance composites correlate highly (.81 for E5 Soldiers, .82 for E6 Soldiers).

Observed Performance Rating Scale Scores with Future Performance

A more detailed look at the relations between observed and future performance can be found in Tables 3.14 and 3.15, which contain correlations between the 18 Observed Performance Rating Scales and the 6 Expected Future Performance Rating Scales. These tables also contain the observed and future performance composites. The covariation pattern is again quite similar across pay grades. Such similarity can be seen by ranking the 18 correlations of each future performance factor with the observed performance scores and then correlating these ranks across pay grade. The resulting rank-order correlations range from .79 to .87, with five of the six values exceeding .80.

The rank-order data also provided insight into the observed performance scores that correlate most highly with expected future performance (i.e., *Leadership Skills*, *Problem Solving*, and *Self-Management* for E5 Soldiers; *MOS-Specific Knowledge*, *Common Task Knowledge and Skill*, and *Leadership Skills* for E6 Soldiers). *Cultural Tolerance* was least correlated with future performance for both pay grades. For E5 Soldiers, *Computer Skills* correlated lowest or next-to-lowest with the future performance factors except for *Use of Technology*, with which it correlated higher than any other dimension of observed performance (as one would expect). *Supporting Peers* also evidenced relatively low correlations with future performance. For E6 Soldiers, *Supporting Peers* and *Soldier Quality of Life* correlated relatively low with the expected future performance scales. The three observed performance scores that had notably higher relationships with expected future performance for E5 Soldiers than for E6 Soldiers were *Level of Effort/Initiative*, *Self-Management*, and *Soldier Quality of Life*. The three observed performance scores that had notably higher relationships with expected future performance for E6 Soldiers than for E5 Soldiers were *MOS-Specific Knowledge*, *Common Task Knowledge*, and *Information Management*.

Table 3.13. Correlations between the Expected Future Performance Rating Scales and the Six Observed Performance Factors (also Includes Composite Scores)

E5 Correlations	TPF: Technical Performance	LDS: Leadership: Structure	EIS: Effort-Integrity-Selfless Service	LDC: Leadership: Consideration	INF: Information Management	SLF: Individual Self-Management	Observed Ratings Composite
EFP: Scenario 1 Self-Direction	.62	.72	.63	.55	.47	.65	.75
EFP: Scenario 2 Use of Technology	.42	.48	.38	.39	.63	.46	.56
EFP: Scenario 3 Scope Technical Skills	.56	.61	.52	.41	.47	.51	.63
EFP: Scenario 4 Broader Leadership	.59	.69	.60	.52	.43	.65	.71
EFP: Scenario 5 Manage Multi Operational Functions	.58	.66	.56	.49	.48	.63	.70
EFP: Scenario 6 Adaptability And Stamina	.54	.65	.55	.46	.30	.59	.64
Expected Future Performance Ratings Composite	.67	.78	.66	.58	.56	.71	.81

E6 Correlations	TPF: Technical Performance	LDS: Leadership: Structure	EIS: Effort-Integrity-Selfless Service	LDC: Leadership: Consideration	INF: Information Management	SLF: Individual Self-Management	Observed Ratings Composite
EFP: Scenario 1 Self-Direction	.70	.74	.61	.46	.51	.62	.76
EFP: Scenario 2 Use of Technology	.50	.50	.38	.34	.68	.36	.58
EFP: Scenario 3 Scope Technical Skills	.60	.66	.52	.46	.56	.54	.70
EFP: Scenario 4 Broader Leadership	.64	.67	.54	.42	.49	.61	.71
EFP: Scenario 5 Manage Multi Operational Functions	.59	.63	.48	.40	.49	.51	.65
EFP: Scenario 6 Adaptability And Stamina	.58	.59	.50	.34	.38	.63	.63
Expected Future Performance Ratings Composite	.74	.77	.62	.49	.63	.67	.82

Note. $n_{E5} = 600$; $n_{E6} = 388$. All correlations significant at $p < .001$.

Table 3.14. Correlations between the Expected Future Performance Rating Scales and the Observed Performance Rating Scales (also Includes Composite Scores): E5 Soldiers

	EFP1	EFP2	EFP3	EFP4	EFP5	EFP6	EFP Composite
OPR: Rating 1 MOS Specific	.56	.35	.52	.53	.54	.49	.61
OPR: Rating 2 Common Task Knowledge & Skill	.56	.41	.48	.53	.49	.48	.60
OPR: Rating 3 Computer Skills	.22	.55	.29	.20	.25	.06	.31
OPR: Rating 4 Writing Skill	.44	.48	.39	.42	.43	.31	.50
OPR: Rating 5 Oral Communication Skill	.49	.39	.45	.48	.48	.45	.56
OPR: Rating 6 Level Of Effort/Initiative	.60	.36	.50	.58	.54	.52	.63
OPR: Rating 7 Adaptability	.57	.37	.45	.49	.49	.51	.59
OPR: Rating 8 Self-Management	.61	.46	.49	.61	.58	.49	.66
OPR: Rating 9 Integrity & Discipline	.51	.31	.41	.48	.44	.45	.53
OPR: Rating 10 Acting As A Role Model	.55	.36	.42	.55	.54	.56	.61
OPR: Rating 11 Supporting Peers	.46	.34	.35	.46	.41	.40	.50
OPR: Rating 12 Cultural Tolerance	.31	.23	.21	.24	.26	.26	.31
OPR: Rating 13 Selfless Service	.51	.32	.42	.47	.46	.44	.54
OPR: Rating 14 Leadership Skills	.67	.39	.53	.66	.61	.59	.71
OPR: Rating 15 Soldier Quality Of Life	.51	.34	.41	.52	.48	.42	.54
OPR: Rating 16 Training Others	.56	.37	.50	.55	.54	.54	.62
OPR: Rating 18 Problem-Solving	.62	.40	.54	.59	.54	.52	.65
OPR: Rating 19 Information Management	.50	.50	.48	.45	.51	.38	.57
Observed Ratings Composite	.75	.56	.63	.71	.70	.64	.81

Note. EFP1 = Self-Direction; EFP2 = Use of Technology; EFP3 = Scope Technical Skills; EFP4 = Broader Leadership; EFP5 = Manage Multi Operational Functions; EFP6 = Adaptability and Stamina. $n = 600$. All correlations significant at $p < .01$.

Construct Validity

To evaluate empirically the construct validity of the performance rating scales, we would need additional criterion measures that tapped the same performance dimensions but employed different methods of measurement (as was possible in Project A; see Knapp, C.H. Campbell, Borman, Pulakos, & Hanson, 2001). The correlations between observed and future performance provide some insight into what factors the supervisors are weighting most heavily when assessing a Soldier's performance in the Army of the future. They do not of themselves, however, allow us to assess the degree to which variation in the scores of each scale stems from the construct targeted for measurement.

Table 3.15. Correlations between the Expected Future Performance Rating Scales and the Observed Performance Rating Scales (also Includes Composite Scores): E6 Soldiers

	EFP1	EFP2	EFP3	EFP4	EFP5	EFP6	EFP Composite
OPR: Rating 1 MOS Specific	.63	.44	.54	.58	.54	.50	.66
OPR: Rating 2 Common Task Knowledge & Skill	.62	.45	.54	.58	.51	.56	.67
OPR: Rating 3 Computer Skills	.32	.62	.41	.31	.31	.23	.44
OPR: Rating 4 Writing Skill	.44	.47	.44	.42	.42	.32	.51
OPR: Rating 5 Oral Communication Skill	.53	.42	.49	.50	.47	.42	.58
OPR: Rating 6 Level Of Effort/Initiative	.59	.32	.48	.54	.44	.48	.58
OPR: Rating 7 Adaptability	.57	.44	.50	.53	.46	.45	.60
OPR: Rating 8 Self-Management	.59	.37	.50	.55	.46	.49	.60
OPR: Rating 9 Integrity & Discipline	.46	.33	.40	.39	.37	.39	.48
OPR: Rating 10 Acting As A Role Model	.49	.27	.44	.51	.42	.59	.56
OPR: Rating 11 Supporting Peers	.43	.31	.40	.35	.33	.27	.43
OPR: Rating 12 Cultural Tolerance	.25	.24	.32	.28	.22	.20	.31
OPR: Rating 13 Selfless Service	.49	.30	.42	.43	.39	.37	.49
OPR: Rating 14 Leadership Skills	.66	.33	.53	.60	.52	.52	.64
OPR: Rating 15 Soldier Quality Of Life	.39	.25	.36	.36	.37	.34	.42
OPR: Rating 16 Training Others	.57	.38	.56	.51	.52	.53	.63
OPR: Rating 18 Problem-Solving	.60	.40	.54	.55	.54	.44	.62
OPR: Rating 19 Information Management	.52	.56	.54	.50	.48	.41	.61
Observed Ratings Composite	.76	.58	.70	.71	.65	.63	.82

Note. EFP1 = Self-Direction; EFP2 = Use of Technology; EFP3 = Scope Technical Skills; EFP4 = Broader Leadership; EFP5 = Manage Multi Operational Functions; EFP6 = Adaptability and Stamina. $n = 388$. All correlations significant at $p < .01$.

Summary

The observed and expected future performance rating scales exhibit satisfactory reliability. The estimates are based on reasonable sample sizes, with most Soldiers being rated by at least one supervisor. A mail-back system ensured maximal data capture and did not reduce the reliability of the ratings—indeed, the reliability of the ratings including the mail-back responses increased slightly for three of the four instrument/grade combinations examined. Correlations between observed and expected future performance were quite similar across pay grades and evidenced sensible covariation patterns. The observed performance scales that correlated most highly with expected future performance in each pay grade differed somewhat, although the *Leadership Skills* scale exhibited high correlations in both grades.

CHAPTER 4: SIMULATED PROMOTION POINT WORKSHEET (SimPPW)

Dan J. Putka and Roy C. Campbell
HumRRO

Overview

The operational Promotion Point Worksheet (PPW) forms the basis of the Army's current NCO promotion system at the E5 and E6 levels. The PPW was simulated to provide a standard against which the validity of other potential predictors could be compared. Our intent was to determine whether alternative predictors (a) were more valid predictors of future NCO performance than the operational PPW, and (b) could offer any incremental validity beyond the operational PPW.

Instrument Description

The simulated PPW (SimPPW) was developed as part of a broader instrument called the Personnel File Form-21 (PFF21). The PFF21 comprises SimPPW content (the focus of this chapter), as well as other content not used in this validation effort but related to Soldiers' experiences. Further details of the development of the PFF21 can be found in Knapp et al. (2002). A copy of the PFF21 is provided in Appendix D.

The operational PPW was the primary source of content for the SimPPW. Soldiers receive promotion points in six areas on the operational PPW: (a) Commander's Evaluation; (b) Promotion Board points; (c) Awards, Certificates, and Military Achievements; (d) Military Education; (e) Civilian Education; and (f) Military Training. Promotion points for the first two areas are awarded by a Soldier's commander and promotion board members at the time a Soldier is up for promotion, whereas points for the latter four areas are allocated by the personnel system based on Soldier records.

Unlike the operational PPW, the SimPPW is a self-report measure designed to capture promotion points awarded in the latter four PPW areas only. Unfortunately, obtaining accurate, timely assessment of Commander's Evaluation and of Promotion Board points via a self-report measure was not feasible for this effort, particularly given our concurrent validation design. As stated above, these points are not awarded to a Soldier until he or she is up for promotion. Thus, any points that Soldiers would have reported in these areas could have potentially come from previous promotions, and may not have accurately reflected the points the Soldier would currently receive in these areas.

Furthermore, in developing the SimPPW, we assumed that Commander's Evaluation and Promotion Board points would not contribute a substantial amount of variation to Soldiers' operational PPW scores. Specifically, Army subject matter experts (SMEs) indicated that these points were often awarded without substantial variation (e.g., on an "all-or-nothing" basis where Soldiers recommended for promotion get the maximum number of points). Hence, their inclusion essentially amounts to adding a constant to each Soldier's total score and thus is unlikely to affect the rank order of Soldiers to any significant degree. As such, our efforts focused on simulating the administrative components of the PPW. These components constitute most of the meaningful

variability, and we were confident that we could obtain good estimates of Soldiers' current promotion points in these areas.

Description of Simulated Scores

SimPPW Awards

The operational PPW gives Soldiers promotion points for obtaining various awards, certificates, and military achievements. Examples of awards and achievements for which a Soldier can receive points include a Combat Infantry Badge, Pathfinder Badge, Special Forces Tab, Distinguished Honor Graduate, and Soldier/NCO of the Quarter-Brigade Level. Although it is unclear how the Army initially assigned points for these awards, more prestigious awards generally are worth more promotion points. A *simulated* PPW Awards score (SimPPW Awards) was calculated for this effort by assigning promotion points to self-reported awards, certificates, and military achievements from the PFF21 (based on operational PPW specifications) and summing these points for each Soldier. SimPPW Award scores were capped at 100 points to mimic operational practice.

SimPPW Military Education

The operational PPW also gives Soldiers promotion points for completing various military education programs. For example, Soldiers can earn promotion points by attending the Primary Leadership Development Course (PLDC), Special Forces Training, Airborne School, and Nuclear, Biological, and Chemical (NBC) School. As with awards and military achievements, educational programs contribute different numbers of points depending, in general, on their levels of prestige. For example, the Special Forces Qualification Course is worth more points than Airborne School. A simulated PPW Military Education score (SimPPW Military Education) was calculated for this effort by assigning promotion points to self-reported military educational experiences from the PFF21 (based on operational PPW specifications)[8] and summing these points for each Soldier. SimPPW Military Education scores were capped at 200 points to be consistent with operational practice.

SimPPW Civilian Education

The operational PPW gives Soldiers promotion points for completing various types of civilian higher education. For example, Soldiers can earn 1.5 promotion points for each semester hour of school they complete (e.g., vocational school, trade school, college) and 10 promotion points for each degree they receive (e.g., associates, bachelors, masters). A simulated PPW Civilian Education score (SimPPW Civilian Education) was calculated for this effort by assigning promotion points to self-reported civilian educational experiences from the PFF21 (based on operational PPW specifications) and summing these points for each Soldier. SimPPW Civilian Education Scores were capped at 100 points (per operational practice).

---

[8] In calculating the simulated Military Education score for this effort, soldiers who attended BNCOC were given 40 points regardless of attendance duration. This change was made to reflect a recent shift in Army policy.

SimPPW Military Training

The operational PPW gives Soldiers promotion points for achieving high levels of marksmanship and physical fitness. For example, Soldiers can earn up to 50 promotion points based on their Army Physical Fitness Test (APFT) scores and up to 50 points based on their last weapons qualification (e.g., expert, sharpshooter). A simulated PPW Military Training score (SimPPW Military Training) was calculated for this effort by (a) assigning promotion points to the self-reported APFT score from the PFF21 (based on operational PPW specifications), (b) assigning promotion points to the self-reported weapons qualification based on an earlier PPW metric (Unqualified = 0 points, Marksman = 10 points, Sharpshooter = 30 points, Expert = 50 points),[9] and (c) summing these points for each Soldier.

SimPPW Composite

A simulated PPW Composite score (SimPPW Composite) was calculated for each Soldier by summing the four simulated scores described above. The maximum score that a Soldier could receive on this composite was 500. The maximum score on the operational PPW is 800. Differences in point totals arise because the simulated PPW does not include Commander's Evaluation points (max 150) or Promotion Board points (max 150).

Results

Data Preparation

Soldiers' responses to items that contributed to SimPPW scores were carefully screened prior to conducting any validation analyses. SimPPW data were first reviewed for outlying responses. Because some PFF21 items asked Soldiers to report on open-ended response scales counts of experiences they had (e.g., number of certificates of achievement, number of semester hours), there was the potential for Soldiers to report unrealistically high values. To mitigate against such unrealistic responses, upper bounds for "permissible" responses on items with open-ended response scales were established based on those used during the field test. For example, E4 Soldiers who reported having more than 15 certificates of achievement (item 3) were assigned a missing response for that item. For E5 and E6 Soldiers, the upper bound for item 3 was raised to 20. In the case of civilian semester hours (item 5), the upper bound of 250 semester hours (across all three education types) was constant across pay grades. Of the 1,890 Soldiers who completed the PFF21, 37 had non-permissible certificate of achievement responses and 8 had non-permissible civilian semester hour responses.

Upon completing the review for outlying responses, we examined the extent of missing data. Based on our goal to maintain sample sizes at high levels, we imputed missing values (including the non-permissible responses identified above) for several items that contributed to the SimPPW scores. Specifically, we imputed missing certificate of achievement counts (item 3), and the sum of the civilian education semester hours (item 5) using the regression-based strategy

[9] A recent change to the operational PPW resulted in a more complicated method for obtaining this score that factors in, for example, the type of weapon used. We used the simpler original formula because of limitations in what we could do with a self-report data collection format.

described in Chapter 2.[10] We imputed missing responses regarding college degrees (item 6), APFT scores (item 9), and weapons qualifications (item 10) using the hot-deck imputation strategy described in Chapter 2 (with pay grade and MOS-type used as cross-classification variables). Of the 1,890 Soldiers who completed the PFF21, only 45 had one or more of their responses imputed (2.3%) using the regression-based strategy. Fifty-one Soldiers (2.7%) had one or more of their responses imputed using the hot-deck imputation strategy.

Relations among Simulated PPW Scores

Simulated PPW score intercorrelations are shown in Table 4.1. For the most part, low to moderate intercorrelations emerged among SimPPW scores. One notable trend was the decreasing correlation between SimPPW Awards and the SimPPW Composite with increases in pay grade (E4: .66, E5: .54, E6: .23). The trend was likely a result of the 100-point cap placed on SimPPW Awards scores. Specifically, a much greater percentage of E6 Soldiers reached the 100-point cap on Awards (93.0%), compared to E5 (45.1%) and E4 Soldiers (4.2%). To the extent that a group of Soldiers achieved the maximum score on Awards, variance in Awards scores is reduced and thus correlations for E6 Soldiers are likely attenuated relative to correlations for E5 and E4 Soldiers.

Table 4.1. Simulated PPW Score Intercorrelations

Predictor	SimPPW Awards	SimPPW Mil Ed	SimPPW Civ Ed	SimPPW Mil Tr
E4 Soldiers				
SimPPW Awards	.			
SimPPW Military Education	.17*	.		
SimPPW Civilian Education	.03	.09*	.	
SimPPW Military Training	.14*	.12*	.00	.
SimPPW Composite	.66*	.59*	.46*	.55*
E5 Soldiers				
SimPPW Awards	.			
SimPPW Military Education	.23*	.		
SimPPW Civilian Education	.11*	.23*	.	
SimPPW Military Training	.08*	.03	-.02	.
SimPPW Composite	.54*	.80*	.59*	.30*
E6 Soldiers				
SimPPW Awards	.			
SimPPW Military Education	.08*	.		
SimPPW Civilian Education	.09*	.13*	.	
SimPPW Military Training	.03	.07*	-.03	.
SimPPW Composite	.23*	.78*	.64*	.32*

Note. $n_{E4} = 448$; $n_{E5} = 885$; $n_{E6} = 555$. Correlations are uncorrected.

*$p < .05$ (one-tailed).

[10] We decided not to impute missing APFT scores (item 9) using the regression-based strategy because we found that no composite of existing PFF21 items provided a high enough *R* value to justify using that composite to predict the missing scores.

Descriptive Statistics

Descriptive statistics for SimPPW scores broken down by subgroup (pay grade, race, gender and CMF cluster) are presented in Tables 4.2 through 4.11. Raw and conditional effect sizes were calculated using the methods described in Chapter 3.

Table 4.2. Subgroup Differences by Pay Grade, Gender, and Race for SimPPW Awards

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	78	33.65	27.57	-0.37	.003	67	31.91	27.66	-0.51	.002
Male	364	43.58	26.76			319	45.62	26.83		
Race										
Black	92	45.04	30.53	0.13	.280	89	41.05	30.10	0.18	.172
White	299	41.51	26.38			297	36.48	26.01		
E5										
Gender										
Female	111	77.64	25.41	-0.16	.123	91	71.21	25.40	-0.34	.044
Male	770	81.44	24.06			676	79.11	23.28		
Race										
Black	246	82.28	23.81	0.08	.300	245	75.28	23.24	0.01	.928
White	523	80.35	24.17			522	75.04	23.65		
E6										
Gender										
Female	58	96.47	14.11	-0.35	.030	47	96.45	15.41	-0.33	.606
Male	496	98.78	6.53			429	98.66	6.76		
Race										
Black	183	97.69	10.89	-0.23	.094	182	96.62	10.99	-0.36	.520
White	297	98.96	5.63			294	98.49	5.27		
Grade										
E5	885	80.83	24.40	1.42	<.001	767	75.16	23.52	1.35	<.001
E4	448	41.87	27.36			386	38.77	26.97		
E6	555	98.54	7.69	0.73	<.001	476	97.56	7.93	0.95	<.001
E5	885	80.83	24.40			767	75.16	23.52		
E6	555	98.54	7.69	2.07	<.001	476	97.56	7.93	2.18	<.001
E4	448	41.87	27.36			386	38.77	26.97		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 4.3. Differences between CMF Clusters for SimPPW Awards

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	69	56	34.25	36.10	28.27	29.66	—	0.09	-0.28	0.22	.	-0.14 *
2. INT	21	20	30.71	38.46	22.87	23.84	-0.13	—	-0.37	0.13	.	-0.23
3. CBO	176	160	42.99	28.42	25.54	25.38	0.32 *	0.45 *	—	0.50 *	.	0.14
4. LOG	143	119	45.64	41.97	29.05	28.46	0.42 **	0.55 **	0.10	—	.	-0.36
5. CPA	12	10	52.75	55.42	26.16	26.39	.	.	.	.	—	.
6. COM	24	21	35.83	32.23	23.99	25.67	0.06	0.19	-0.26	-0.36	.	—
Overall	448		41.87		27.36							
E5 Soldiers												
1. ADM	85	70	83.41	83.97	23.44	21.61	—	-0.37	-0.66	-0.12	-0.43 *	-0.67 **
2. INT	37	35	69.00	75.35	31.48	30.96	-0.59 **	—	-0.30	0.24	-0.06	-0.30
3. CBO	332	287	81.45	68.33	22.95	22.24	-0.08	0.51 *	—	0.54	0.24	-0.00
4. LOG	290	249	83.94	81.12	23.81	22.97	0.02	0.61 **	0.10	—	-0.31	-0.54 **
5. CPA	80	74	75.74	73.92	25.39	26.19	-0.31 *	0.28	-0.23 *	-0.34 **	—	-0.24
6. COM	59	52	72.90	68.28	26.07	26.19	-0.43 **	0.16	-0.35 **	-0.45 **	-0.12	—
Overall	885		80.83		24.40							
E6 Soldiers												
1. ADM	60	51	98.08	98.60	7.08	7.36	—	.	-1.37	-0.29	0.10	-0.01
2. INT	21	16	95.62	104.8	14.41	16.95	-0.32	—	.	.	.	.
3. CBO	210	183	99.01	87.69	4.58	4.15	0.12	0.44	—	1.08	1.47	1.36
4. LOG	168	145	98.20	96.31	10.50	11.03	0.01	0.34	-0.11	—	0.39	0.28
5. CPA	65	55	98.92	99.38	5.83	5.51	0.11	0.43	-0.01	0.09	—	-0.11
6. COM	31	26	99.19	98.53	3.67	4.08	0.14	0.47	0.02	0.13	0.04	—
Overall	555		98.54		7.69							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 4.4. Subgroup Differences by Pay Grade, Gender, and Race for SimPPW Military Education

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	78	15.95	17.92	0.03	.786	67	9.34	16.86	-0.40	.251
Male	364	15.2	22.99			319	18.69	23.22		
Race										
Black	92	17.97	31.47	0.17	.227	89	14.26	32.47	0.03	.937
White	299	14.73	18.86			297	13.77	18.40		
E5										
Gender										
Female	111	68.18	51.82	0.14	.190	91	56.55	49.68	-0.26	.127
Male	770	62.11	44.58			676	67.75	43.02		
Race										
Black	246	73.81	52.25	0.36	<.001	245	65.44	50.44	0.16	.178
White	523	58.85	42.06			522	58.87	40.38		
E6										
Gender										
Female	58	116.86	47.33	-0.11	.416	47	100.50	48.18	-0.44	.119
Male	496	121.97	44.92			429	119.07	42.44		
Race										
Black	183	117.07	45.26	-0.15	.109	182	105.55	44.10	-0.20	.203
White	297	123.79	44.18			294	114.02	42.30		
Grade										
E5	885	63.09	45.71	2.17	<.001	767	62.15	43.81	2.16	<.001
E4	448	15.19	22.07			386	14.01	22.31		
E6	555	121.32	45.21	1.27	<.001	476	109.79	42.98	1.09	<.001
E5	885	63.09	45.71			767	62.15	43.81		
E6	555	121.32	45.21	4.81	<.001	476	109.79	42.98	4.29	<.001
E4	448	15.19	22.07			386	14.01	22.31		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 4.5. Differences between CMF Clusters for SimPPW Military Education

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	69	56	20.78	21.50	24.70	25.30	—	0.03	-0.78	-0.31	.	-0.87
2. INT	21	20	13.00	22.26	16.16	16.77	-0.35	—	-0.82	-0.34	.	-0.90
3. CBO	176	160	13.96	4.05	18.05	18.55	-0.31	0.04	—	0.48	.	-0.09
4. LOG	143	119	14.49	14.68	25.57	26.82	-0.28	0.07	0.02	—	.	-0.56
5. CPA	12	10	34.25	19.45	28.53	29.55	.	.	.	.	—	.
6. COM	24	21	5.13	2.14	10.29	8.23	-0.71 **	-0.36	-0.40 *	-0.42	.	—
Overall	448		15.19		22.07							
E5 Soldiers												
1. ADM	85	70	93.92	94.64	62.54	61.16	—	-1.01 **	-1.18 **	-0.65 **	-0.67 **	-0.94 **
2. INT	37	35	44.59	50.43	22.79	15.87	-1.08 **	—	-0.17	0.36	0.34	0.07
3. CBO	332	287	51.80	42.95	32.93	34.16	-0.92 **	0.16	—	0.53	0.51	0.24
4. LOG	290	249	67.13	66.26	47.60	48.02	-0.59 **	0.49 **	0.34 **	—	-0.03	-0.29
5. CPA	80	74	75.90	65.15	54.24	52.92	-0.39	0.68 **	0.53 **	0.19	—	-0.27
6. COM	59	52	56.00	53.50	39.92	41.70	-0.83 **	0.25 *	0.09	-0.24	-0.44 *	—
Overall	885		63.09		45.71							
E6 Soldiers												
1. ADM	60	51	132.13	131.97	49.84	49.12	—	.	-0.80 **	-0.32 *	-0.20	-0.52 *
2. INT	21	16	92.10	78.15	39.01	27.80	-0.89 **	—	.	.	.	.
3. CBO	210	183	114.13	97.55	40.63	38.40	-0.40 *	0.49 *	—	0.48	0.60	0.28
4. LOG	168	145	123.18	118.12	47.49	47.44	-0.20	0.69 **	0.20 *	—	0.12	-0.20
5. CPA	65	55	141.12	123.29	43.79	43.95	0.20	1.08 **	0.60 **	0.40 *	—	-0.32
6. COM	31	26	117.23	109.63	40.15	40.33	-0.33	0.56 *	0.07	-0.13	-0.53 *	—
Overall	555		121.32		45.21							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 4.6. Subgroup Differences by Pay Grade, Gender, and Race for SimPPW Civilian Education

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	78	10.83	20.87	0.12	.321	67	-0.04	12.38	-0.50	.090
Male	364	8.19	21.45			319	10.11	20.50		
Race										
Black	92	8.91	19.72	0.09	.432	89	5.01	19.80	0.00	.991
White	299	7.10	19.16			297	5.06	19.30		
E5										
Gender										
Female	111	35.62	36.25	0.52	<.001	91	25.67	35.04	0.15	.428
Male	770	20.51	29.17			676	21.40	29.22		
Race										
Black	246	26.36	31.26	0.18	.022	245	23.05	29.01	-0.03	.789
White	523	20.89	30.56			522	24.01	30.33		
E6										
Gender										
Female	58	66.32	37.03	0.22	.118	47	44.93	33.48	-0.42	.016
Male	496	58.46	36.02			429	59.10	33.76		
Race										
Black	183	64.35	35.73	0.27	.004	182	54.55	32.45	0.15	.204
White	297	54.57	35.84			294	49.48	34.49		
Grade										
E5	885	22.40	30.50	0.62	<.001	767	23.53	29.92	0.95	<.001
E4	448	8.85	21.69			386	5.03	19.42		
E6	555	59.18	36.23	1.21	<.001	476	52.02	33.73	0.95	<.001
E5	885	22.40	30.50			767	23.53	29.92		
E6	555	59.18	36.23	2.32	<.001	476	52.02	33.73	2.42	<.001
E4	448	8.85	21.69			386	5.03	19.42		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 4.7. Differences between CMF Clusters for SimPPW Civilian Education

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	69	56	10.55	8.53	22.04	17.44	—	-0.11	-0.74	-0.13	.	0.05
2. INT	21	20	14.00	6.34	21.56	20.70	0.16	—	-0.63	-0.01	.	0.16
3. CBO	176	160	6.62	5.98	19.73	19.27	-0.18	-0.34	—	0.62	.	0.79
4. LOG	143	119	8.55	6.08	20.70	16.62	-0.09	-0.25	0.09	—	.	0.18
5. CPA	12	10	12.33	5.73	28.02	32.48	.	.	.	.	—	.
6. COM	24	21	16.71	9.52	34.16	30.43	0.28	0.12	0.46 *	0.38	.	—
Overall	448		8.85		21.69							
E5 Soldiers												
1. ADM	85	70	34.89	38.08	35.83	33.48	—	-0.74 **	-0.92 *	-0.28 *	-0.36	-0.61 **
2. INT	37	35	16.38	15.90	27.73	28.81	-0.61 **	—	-0.18	0.46	0.38	0.13
3. CBO	332	287	15.27	10.44	25.31	26.10	-0.64 **	-0.04	—	0.64	0.56	0.31
4. LOG	290	249	26.96	29.62	32.67	32.53	-0.26	0.35 *	0.38 **	—	-0.08	-0.33
5. CPA	80	74	27.77	27.30	31.26	32.41	-0.23	0.37 *	0.41 **	0.03	—	-0.25
6. COM	59	52	19.36	19.86	28.66	29.03	-0.51 *	0.10	0.13	-0.25	-0.28	—
Overall	885		22.40		30.50							
E6 Soldiers												
1. ADM	60	51	80.68	78.19	30.49	32.31	—	.	-1.36 **	-0.41 **	-0.57 **	-0.95 **
2. INT	21	16	42.10	32.48	39.14	41.14	-1.06 **	—	.	.	.	.
3. CBO	210	183	45.63	32.26	34.53	34.26	-0.97 **	0.10	—	0.95 **	0.79 *	0.41
4. LOG	168	145	67.79	64.21	34.98	33.69	-0.36 **	0.71 **	0.61 **	—	-0.16	-0.54 **
5. CPA	65	55	68.45	58.95	33.11	33.48	-0.34 *	0.73 **	0.63 **	0.02	—	-0.38
6. COM	31	26	54.77	46.01	29.87	27.62	-0.71 **	0.35	0.25	-0.36	-0.38	—
Overall	555		59.18		36.23							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 4.8. Subgroup Differences by Pay Grade, Gender, and Race for SimPPW Military Training

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	78	43.12	17.38	-0.76	<.001	67	45.01	17.14	-0.49	.007
Male	364	60.44	22.88			319	55.57	21.50		
Race										
Black	92	51.09	22.07	-0.34	.004	89	48.43	20.14	-0.18	.212
White	299	58.93	23.05			297	52.15	21.05		
E5										
Gender										
Female	111	56.41	19.16	-0.68	<.001	91	57.70	20.54	-0.55	.004
Male	770	69.26	18.85			676	67.75	18.17		
Race										
Black	246	66.02	19.52	-0.12	.117	245	62.44	18.83	-0.03	.807
White	523	68.37	19.27			522	63.01	18.26		
E6										
Gender										
Female	58	57.43	23.62	-0.78	<.001	47	56.78	25.18	-0.69	.002
Male	496	71.29	17.85			429	68.45	16.90		
Race										
Black	183	68.87	20.14	-0.06	.563	182	62.82	18.71	0.02	.872
White	297	69.90	18.12			294	62.40	17.22		
Grade										
E5	885	67.68	19.36	0.45	<.001	767	62.72	18.44	0.60	<.001
E4	448	57.31	22.94			386	50.29	20.85		
E6	555	69.87	18.99	0.11	.060	476	62.61	17.80	-0.01	.957
E5	885	67.68	19.36			767	62.72	18.44		
E6	555	69.87	18.99	0.55	<.001	476	62.61	17.80	0.59	<.001
E4	448	57.31	22.94			386	50.29	20.85		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 4.9. Differences between CMF Clusters for SimPPW Military Training

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	69	56	48.51	47.12	20.42	19.01	—	0.23	0.56	-0.04	.	-0.28
2. INT	21	20	54.43	51.92	20.18	19.48	0.26	—	0.33	-0.28	.	-0.51
3. CBO	176	160	69.18	58.81	22.36	22.63	0.90 **	0.64 **	—	-0.61	.	-0.84 *
4. LOG	143	119	49.35	46.19	19.41	19.43	0.04	-0.22	-0.86 **	—	.	-0.23
5. CPA	12	10	59.75	56.40	20.92	24.18	.	.	.	.	—	.
6. COM	24	21	45.29	41.31	20.05	17.98	-0.14	-0.40	-1.04 **	-0.18	.	—
Overall	448		57.31		22.94							
E5 Soldiers												
1. ADM	85	70	62.79	63.00	20.04	20.24	—	-0.11	0.17	-0.22	0.14	-0.07
2. INT	37	35	63.35	61.05	17.68	17.33	0.03	—	0.27	-0.12	0.25	0.04
3. CBO	332	287	74.51	66.11	17.46	17.37	0.60 **	0.58 **	—	-0.39	-0.03	-0.24
4. LOG	290	249	62.54	58.86	18.97	18.63	-0.01	-0.04	-0.62 **	—	0.37	0.15
5. CPA	80	74	69.06	65.63	17.40	17.37	0.32 *	0.29	-0.28 *	0.34 **	—	-0.21
6. COM	59	52	62.83	61.70	22.74	22.73	0.00	-0.03	-0.60 **	0.01	-0.32 *	—
Overall	885		67.68		19.36							
E6 Soldiers												
1. ADM	60	51	65.38	66.13	22.38	22.15	—	.	0.05	-0.32	0.04	-0.70 *
2. INT	21	16	67.10	61.62	19.50	21.23	0.09	—	.	.	.	.
3. CBO	210	183	76.20	67.05	15.20	15.13	0.57 **	0.48 *	—	-0.37	-0.01	-0.76
4. LOG	168	145	65.79	60.51	20.75	19.77	0.02	-0.07	-0.55 **	—	0.35	-0.39
5. CPA	65	55	70.52	66.79	15.01	13.50	0.27	0.18	-0.30 **	0.25	—	-0.74 *
6. COM	31	26	58.23	53.57	19.28	20.54	-0.38	-0.47	-0.95 **	-0.40	-0.65 **	—
Overall	555		69.87		18.99							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 4.10. Subgroup Differences by Pay Grade, Gender, and Race for SimPPW Composite

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	78	103.55	52.23	-0.45	<.001	67	86.22	48.18	-0.84	.001
Male	364	127.40	52.50			319	129.99	52.18		
Race										
Black	92	123.01	62.01	0.01	.906	89	108.75	60.19	0.03	.900
White	299	122.26	50.04			297	107.46	48.80		
E5										
Gender										
Female	111	237.86	82.75	0.06	.544	91	211.14	79.14	-0.35	.038
Male	770	233.32	72.13			676	236.01	70.92		
Race										
Black	246	248.47	81.33	0.28	<.001	245	226.21	78.68	0.08	.507
White	523	228.46	70.51			522	220.93	68.51		
E6										
Gender										
Female	58	337.08	78.39	-0.21	.144	47	298.66	70.44	-0.77	<.001
Male	496	350.50	64.61			429	345.28	60.24		
Race										
Black	183	347.99	70.59	0.01	.903	182	319.55	65.21	-0.08	.584
White	297	347.22	63.38			294	324.39	58.67		
Grade										
E5	885	234.00	73.69	2.07	<.001	767	223.57	71.88	2.24	<.001
E4	448	123.21	53.40			386	108.11	51.55		
E6	555	348.90	66.33	1.56	<.001	476	321.97	61.22	1.37	<.001
E5	885	234.00	73.69			767	223.57	71.88		
E6	555	348.90	66.33	4.23	<.001	476	321.97	61.22	4.15	<.001
E4	448	123.21	53.40			386	108.11	51.55		

Note. Effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 4.11. Differences between CMF Clusters for SimPPW Composite

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	69	56	114.09	113.25	55.50	54.70	—	0.11	-0.54	-0.08	.	-0.54
2. INT	21	20	112.14	118.98	45.87	44.98	-0.04	—	-0.65	-0.20	.	-0.65
3. CBO	176	160	132.76	85.30	49.43	49.37	0.35 *	0.39	—	0.46	.	-0.00
4. LOG	143	119	118.02	108.91	54.86	53.09	0.07	0.11	-0.28 **	—	.	-0.46
5. CPA	12	10	159.08	137.00	70.18	79.10	.	.	.	.	—	.
6. COM	24	21	102.96	85.19	50.27	40.30	-0.21	-0.17	-0.56 **	-0.28	.	—
Overall	448		123.21		53.40							
E5 Soldiers												
1. ADM	85	70	275.01	279.69	82.17	80.99	—	-1.07 **	-1.28 **	-0.61 **	-0.66 **	-1.06 **
2. INT	37	35	193.32	202.73	51.78	45.19	-1.11 **	—	-0.21	0.46	0.41	0.01
3. CBO	332	287	223.02	187.82	58.33	59.06	-0.71 **	0.40 **	—	0.67 *	0.61	0.22
4. LOG	290	249	240.57	235.85	79.27	80.63	-0.47 **	0.64 **	0.24 **	—	-0.05	-0.45 *
5. CPA	80	74	248.47	232.00	81.96	82.35	-0.36 *	0.75 **	0.35 **	0.11	—	-0.40
6. COM	59	52	211.08	203.35	77.49	78.12	-0.87 **	0.24	-0.16	-0.40 **	-0.51 **	—
Overall	885		234.00		73.69							
E6 Soldiers												
1. ADM	60	51	376.28	374.89	64.78	67.07	—	.	-1.48 **	-0.58 **	-0.43	-1.10 **
2. INT	21	16	296.90	277.09	68.80	58.29	-1.20 **	—	.	.	.	.
3. CBO	210	183	334.98	284.55	59.30	56.73	-0.62 **	0.57 *	—	0.89 *	1.04 *	0.38
4. LOG	168	145	354.96	339.15	71.31	68.48	-0.32 *	0.88 **	0.30 **	—	0.15	-0.51 *
5. CPA	65	55	379.02	348.40	55.28	52.37	0.04	1.24 **	0.66 **	0.36 *	—	-0.66 *
6. COM	31	26	329.42	307.74	56.68	55.75	-0.71 **	0.49	-0.08	-0.39	-0.75 **	—
Overall	555		348.90		66.33							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Given the number of effect sizes presented in Tables 4.2 through 4.11, only a few notable findings are summarized here. First, as expected, there were sizable differences in means for SimPPW Awards, SimPPW Military Education, and the SimPPW Composite across pay grades. Such findings support the validity of these scores as measures of Soldiers' military experience. Second, although there was a high level of range restriction in E6 SimPPW Awards scores (recall, 93.0% of E6 Soldiers scored at the upper bound), range restriction appeared to be far less of an issue for E6 Soldiers on the SimPPW Composite. Third, moderate to large gender differences in SimPPW Composite scores were found even after controlling for race, CMF cluster, and pay grade. Specifically, women tended to have scores that were 0.35 (E5 Soldiers) to 0.77 (E6 Soldiers) standard deviation lower than men on the SimPPW Composite (holding race and CMF cluster constant). Lastly, there were some sizable CMF cluster differences in SimPPW Composite scores. Specifically, E5 and E6 Soldiers in the CMF Administration cluster tended to have SimPPW Composite scores that were 0.43 to 1.48 standard deviations higher than Soldiers in the other CMF clusters (holding race and sex constant).

Validity Estimates

Evidence for criterion-related validity was examined by computing zero-order correlations between the SimPPW scores and four criterion scores described in Chapter 3 (i.e., Observed Performance Rating Scales composite, Expected Future Performance Rating Scales composite, Senior NCO Potential Rating, Overall Effectiveness Rating). Separate correlations were computed for E5 and E6 Soldiers, and differences between corresponding correlations (across pay grades) were tested for statistical significance. All correlations were corrected for unreliability in the criterion (using reliability estimates presented in Chapter 3) and direct range restriction on the predictor (using Thorndike's [1949] correction formula). Corrected and raw correlations are presented in Table 4.12. Because the primary focus of this chapter is on formulating a SimPPW Composite score for each Soldier, our discussion of validity will primarily focus on the SimPPW Composite.

The SimPPW Composite showed low to moderate validity for predicting both observed and expected future performance among E5 (.19 for observed performance, .13 for expected performance) and E6 (.13 for observed performance, .18 for expected performance) Soldiers. No significant E5-E6 differences were observed between corresponding correlations involving the SimPPW composite. A similar pattern of estimated validities was obtained for predicting the single-item criteria (Senior NCO Potential Rating and Overall Effectiveness Rating).

One interesting finding regarded the validity estimates for SimPPW Military Training. These estimates tended to be higher than the validity estimates of other SimPPW components, including Military Education (which in operational use is allocated twice as many points as the other components). This trend was more pronounced for E5 Soldiers than for E6 Soldiers but held up across all criteria.

Differential Prediction Analyses

An important aspect of any validation effort is to investigate potential bias in one's measure. The model of bias used in this validation effort is based on Cleary's (1968) model, which recognizes two potential types of bias (intercept and slope bias). The extent of each bias can be estimated by fitting a moderated multiple regression (MMR) model to the data.

Table 4.12. Corrected and Raw Correlations between Simulated PPW Scores and Criteria for E5 and E6 Soldiers

Criterion	Predictor				
	SimPPW Awards	SimPPW Mil Ed	SimPPW Civ Ed	SimPPW Mil Tr	SimPPW Composite
E5 Soldiers					
Observed Performance Composite	.05 (.03)	.12 (.17$_a$*)	.07 (.07*)	.31 (.19$_a$*)	.19 (.19*)
Expected Future Performance Composite	-.04 (-.02)	.08 (.10*)	.02 (.02)	.34 (.18*)	.13 (.11*)
Senior NCO Potential Rating	.04 (.02)	.08 (.12*)	.04 (.04)	.32 (.19*)	.15 (.14*)
Overall Effectiveness Rating	.03 (.02)	.10 (.13*)	.01 (.01)	.35 (.20*)	.15 (.14*)
E6 Soldiers					
Observed Performance Composite	.06 (.01)	.04 (.03)	.09 (.09*)	.08 (.06)	.13 (.09*)
Expected Future Performance Composite	.09 (.02)	.06 (.04)	.07 (.06)	.24 (.16*)	.18 (.11*)
Senior NCO Potential Rating	-.02 (.00)	.12 (.08)	.08 (.07)	.12 (.08)	.18 (.12*)
Overall Effectiveness Rating	-.07 (-.02)	.05 (.04)	.06 (.05)	.15 (.11*)	.13 (.08*)

Note. n_{E5} = 608-613; n_{E6} = 391-397. Correlations corrected for criterion unreliability and for direct range restriction on the predictor appear outside of parentheses. Raw correlations appear inside parentheses. The "a" subscripts on E5 correlations indicate that corresponding E5 and E6 correlations were significantly different from each other, $p < .05$ (two-tailed).

* $p < .05$ (one-tailed).

Intercept bias reflects differences in the intercept terms of regression lines fitted for each subgroup. In the context of MMR analysis, this is evidenced by a significant main effect for subgroup membership (e.g., gender, race). Intercept bias suggests that the instrument would underpredict performance for one group relative to another if a common regression line was used to predict performance.

Slope bias reflects differences in the slopes associated with the instrument in regression lines fit for each subgroup separately (i.e., differential prediction). In the context of MMR analysis, this is evidenced by a significant slope for the interaction between the instrument and subgroup membership. Slope bias suggests that the instrument is more predictive of performance for one subgroup than another.

Table 4.13 presents the results of differential prediction analyses for SimPPW scores by pay grade and criterion, examining gender and race as the demographic variables of interest.[11] Values reported under the "Demographic Main Effect" column are the unstandardized regression weights (*b*) associated with the demographic variable from MMR analyses. These values reflect the predicted difference between subgroups' (females-males, blacks-whites) raw criterion scores at the mean SimPPW score (across subgroups, within pay grade). Values reported under the "SimPPW Score Main Effect" column reflect the predicted change in raw

[11] All SimPPW scores were standardized within pay grade prior to conducting these MMR analyses to ease interpretation of the unstandardized regression weights. The demographic variables were coded as follows for purposes of analysis: race (white = 0, black = 1), gender (male = 0, female = 1).

criterion scores associated with a 1.0 standard deviation increase on the SimPPW score for the given subgroup. For referent groups (e.g., males and whites), these values are simply the unstandardized regression weights associated with the SimPPW score of interest. For the non-referent groups (e.g., females and blacks), these values are the sum of the unstandardized regression weights associated with the SimPPW score of interest, and the cross-product term (SimPPW score x demographic variable). Values under the "r" column reflect uncorrected zero-order correlations between SimPPW scores and criteria for each subgroup separately.

Table 4.13. Differential Prediction Analyses for Simulated PPW Scores

Criterion/Predictor	Demographic Main Effect		SimPPW Score Main Effect				*r*			
			Gender		Race		Gender		Race	
	Gender	Race	M	F	W	B	M	F	W	B
			Observed Performance Composite							
E5 Soldiers										
SimPPW Awards	-.15	-.01	.01	.15	-.02	.11	.01	.17	-.02	.13
SimPPW Military Education	-.18	-.03	.14	.22	.16	.18	**.16**	**.25**	**.17**	**.22**
SimPPW Civilian Education	-.19	-.02	.05	.14	.03	.12	.06	.17	.03	**.15**
SimPPW Military Training	-.12	-.02	.17	.06	.18	.03	**.20**	.07	**.21**	.04
SimPPW Composite	-.16	-.04	.15	.25	.14	.18	**.18**	**.30**	**.17**	**.24**
E6 Soldiers										
SimPPW Awards	.13	-.11	.05	.21	.03	.01	-.01	.17	.03	.01
SimPPW Military Education	-.06	-.10	.01	.17	-.01	.06	.01	.18	-.01	.06
SimPPW Civilian Education	-.12	-.13	.07	.04	.10	.06	**.09**	.04	**.14**	.07
SimPPW Military Training	.08	-.10	.01	.21	.05	.07	.01	**.30**	.07	.09
SimPPW Composite	.00	-.10	.05	.21	.07	.09	.06	.25	.09	.11
			Expected Future Performance Composite							
E5 Soldiers										
SimPPW Awards	-.37*	.01	-.03	.00	-.07	.04	-.03	.00	-.07	.05
SimPPW Military Education	-.38*	.00	.10	.12	.14	.07	**.10**	.13	**.14**	.08
SimPPW Civilian Education	-.39*	.01	.02	.07	.01	.00	.03	.08	.01	.01
SimPPW Military Training	-.34*	-.01	.18	.05	$.22_a$	-.04	**.18**	.05	**.23**	-.04
SimPPW Composite	-.37*	.00	.10	.12	.12	.05	**.11**	.13	**.12**	.06
E6 Soldiers										
SimPPW Awards	-.41*	-.12	.01	.09	.03	.00	.01	.06	.02	.00
SimPPW Military Education	-.30	-.11	.00	.34	-.02	.11	.00	.24	-.02	.10
SimPPW Civilian Education	-.41*	-.14	.06	.03	.05	.11	.07	.03	.06	.10
SimPPW Military Training	-.15	-.10	.11	.27	.15	.19	**.11**	.26	**.16**	**.18**
SimPPW Composite	-.24	-.11	.06	.31	.06	.19	.07	.25	.07	**.17**

Note. Regression analysis sample sizes: $n_{E5\ Gender}$ = 606-611; $n_{E5\ Race}$ = 525-528; $n_{E6\ Gender}$ = 390-396; $n_{E6\ Race}$ = 341-346. Smaller sample sizes underlie the reported correlations because they were calculated for each subgroup separately. The "a" subscripts on the SimPPW main effect values indicate that the SimPPW-by-demographic interaction term was statistically significant, $p < .05$ (two-tailed). Subscripts are located on the subgroup with the higher value. Bolded correlations are statistically significant, $p < .05$ (one-tailed).

*$p < .05$ (two-tailed).

Overall, the results provide little evidence of differential prediction (i.e., slope bias). The only case where differential prediction appeared evident was when using SimPPW Military Training as a predictor of expected future performance for E5 Soldiers. Specifically, the SimPPW Military Training score was more predictive of expected future performance for white E5 Soldiers ($b = 0.22$) than for black E5 Soldiers ($b = -0.04$). Evidence of intercept bias emerged only for gender-based comparisons when expected future performance was the criterion (particularly among E5 Soldiers). Specifically, female E5 Soldiers tended to have scores that were roughly 0.34 to 0.39 lower than males E5 Soldiers (at mean levels of SimPPW scores). These findings suggest that the SimPPW would tend to *overpredict* females' expected future performance if a common regression equation were used.

Summary

The SimPPW Composite score showed low to moderate levels of validity for predicting both current and expected future NCO performance among E5 and E6 Soldiers. Of the SimPPW component scores, SimPPW Military Training appeared to be most predictive of the performance criteria, particularly for E5 Soldiers.

As discussed in Chapter 9, the concurrent design used in this validation effort may unduly affect the validity of experience-based predictors such as the SimPPW. Specifically, based on this design, it is difficult to accurately discern the relationship between Soldiers' SimPPW scores recorded immediately prior to promotion to the next grade, and their performance at that next grade. For example, the sample of Soldiers examined in this effort spanned a wide range of experience levels within grade (e.g., some who were promotion-eligible, and others who were not). Thus, the validity of the SimPPW observed here may more reflect the constraints of the concurrent design, relative to the validity of other predictors that are generally unrelated to Soldiering experience (e.g., temperament, cognitive ability).

Subgroup analyses revealed that women tended to have lower SimPPW composite scores than men, even after controlling for race, CMF, and pay grade differences. Moreover, these analyses revealed that E5/E6 Soldiers in the CMF Administration cluster tended to have significantly higher SimPPW composite scores than E5/E6 Soldiers in other CMF clusters. Again, these differences were sizable even after controlling for other demographic variables (i.e., race, pay grade).

Overall, SimPPW scores did not appear to be differentially predictive for comparisons based on gender and race. However, there was evidence of intercept bias for gender (females' performance being overpredicted) when expected future performance was used as the criterion, particularly among E5 Soldiers.

CHAPTER 5: EXPERIENCE AND ACTIVITIES RECORD (EXACT)

Dan J. Putka
HumRRO

Overview

This chapter describes the validation of a self-report measure designed to capture information about Soldiers' work experiences, activities, and accomplishments indicative of KSAs considered relevant to the performance of 21st-century NCOs. The initiative to include an assessment of experiences for the NCO21 validation effort stems in part from the previous success of similar measures in Project A for predicting job performance of entry-level Soldiers (J. Campbell & Knapp, 2001). Multiple self-report instruments were developed during Project A to capture biodata (e.g., Assessment of Background and Life Experiences), archival information (e.g., Personnel File Form), and Soldier experiences (e.g., Supervisory Experience Questionnaire). In that project, these instruments provided information that predicted Soldier performance.

Instrument Description

The content of the ExAct reflects specific activities and experiences that are not typically documented but may predict performance at the next pay grade. It is a reasonable presumption that Soldiers who have engaged in more of these activities and have done so more often will perform at a higher level than will those with less experience. That is, knowledge of a Soldier's prior experiences should provide useful information for assessing his or her preparedness to perform similar activities in the future.

Forty-six items constitute the validation version of the ExAct. Item writers targeted many KSAs during the course of instrument development: (a) *Writing Skill*; (b) *Computer Skill*; (c) *Motivating Leading, and Supporting Individual Subordinates*; (d) *Directing, Monitoring, and Supervising Individual Subordinates*; (e) *Training Others*; (f) *Team Leadership*; and (g) *Level of Effort and Initiative on the Job* (see Knapp et al., 2002, Chapter 4, for complete details on the development of the ExAct). A copy of the ExAct is presented in Appendix E.

Results

Data Preparation

Soldiers' responses to ExAct items were carefully screened prior to conducting any validation analyses. Of primary concern were missing responses and evidence of patterned responding (e.g., a Soldier responds to every item using the highest point on the scale). Based on a careful review of the data by two members of the NCO21 project team, no Soldier's data were removed for reasons of patterned responding. We then examined missing responses in the data set. To maintain sample sizes at high levels for purposes of validation, we retained any individual who responded to at least 90% of the ExAct items (42 out of 46). Of the 1,893 Soldiers who completed the ExAct, only 11 responded to fewer than 42 items. These 11 Soldiers were eliminated from all further ExAct analyses. Missing item responses for Soldiers that

remained in the sample were imputed using the regression-based method described in Chapter 2. Overall, less than 1% of ExAct data points were imputed (504 of 86,480).[12]

Score Development

Because biographical items typically reflect multiple KSAs (in varying degrees), a total score for such an instrument is often used. A total score is inappropriate, however, if specific items clearly define relatively independent dimensions. In the field test investigation of the ExAct, principal components analyses (PCA) with orthogonal rotation indicated that a two-component structure (reflecting Computer Experience and General Experience) best described the data.

Given the findings of the field test, a confirmatory factor analysis (CFA) was conducted to determine whether the validation data yielded a two-factor structure. Prior to investigating the structure underlying the ExAct data, all items were standardized across the entire sample to place them on the same metric (M=0, SD=1). A CFA was then conducted across all pay grades sampled. In specifying the CFA model, the correlations among factors were constrained to 0 to parallel the orthogonal rotation from the PCA in the field test. Results of the CFA analyses suggested a reasonable fit for the two-factor solution ($\chi^2_{(989)} = 4{,}985.96, p < .001$; $CFI = .96$; $RMSEA = .046$).[13]

Upon closer inspection of the ExAct and the CFA results, we hypothesized that a third factor (reflecting supervisory experience) might be present. Because one of our goals in developing the ExAct was to distinguish more finely between different aspects of experience, we conducted exploratory factor analysis (EFA) within each pay grade, as well as across all pay grades sampled (E4, E5, and E6). All EFA were based on the principle axis factoring extraction method and employed oblique rotation, thus allowing potential factors to covary.[14] In an initial round of EFAs, no set number of factors was specified for purposes of extraction.

Eigenvalues from these initial EFA suggested a three-factor structure. Thus, a set of follow-up EFA constrained the solution to three factors. Table 5.1 presents the pattern matrix resulting from the follow-up EFA on the overall sample.[15] The primary difference between the two- and three-factor solutions is that the second factor from the field test (General Experience) split into two factors (General and Supervisory Experience).

---

[12] The total number of ExAct data points (86,480) is the number of ExAct items (46) times the number of respondents (1,882).

[13] A CFA model where correlations among factors were free to vary was also fitted to the data. Although results revealed a statistically significant improvement in fit compared to the constrained-phi covariance model ($\chi^2_{(1)} = 12.19, p < .01$), these differences did not appear to be meaningful, as other indicators of fit remained very similar (e.g., $CFI = .96$; $RMSEA = .046$). All CFA estimates were based on generalized least squares estimation.

[14] We chose to use EFA (as opposed to PCA) for this effort, because unlike the field test, we no longer had notions of using criterion-reference scoring for the ExAct. The focus of this validation effort was identifying experience-based constructs that comprise the ExAct.

[15] Results of the EFA on the overall sample only are presented because EFA conducted by pay grade revealed very similar factor structures. A follow-up CFA was also conducted on the overall sample. The results of this analysis suggested that the unconstrained three-factor model (where factor correlations were free to vary) provided a similar fit to the data ($\chi^2_{(986)} = 4786.23, p < .001$; $CFI = .96$; $RMSEA = .045$) relative to the unconstrained two-factor model. Because EFA conducted by pay grade revealed very similar factor structures, no CFA were conducted by pay grade.

Table 5.1. ExAct Pattern Matrix: Three-Factor Solution

ExAct Item	Factor		
	General	Computer	Supervisory
34. Conducted primary marksmanship instruction (PMI)	**.69**	-.05	.12
36. Issued a 5 paragraph oral operations order	**.69**	.01	.02
35. Received and implemented a written operations order	**.66**	.02	.06
37. Prepared and submitted a written report of recognition for a subordinate	**.66**	.05	.10
39. Prepared a written plan/schedule of future subordinate activities covering 5 or more days	**.61**	.14	.06
33. Participated as a team leader or above in a live fire exercise (LFX)	**.57**	-.21	.06
38. Prepared and conducted a briefing for 2 or more officer, senior NCO, or civilian personnel	**.56**	.25	-.03
22. Total time spent in a leadership or supervisory position	**.56**	.02	.32
23. Total time spent in MTOE slot assignment	**.51**	.18	.10
31. Served as an assistant instructor in a class of 10 or more people	**.49**	.04	.17
45. Served as a VIP escort	**.47**	.13	-.12
30. Taught a platform class to 5 or more people	**.45**	-.01	.23
42. Conducted an inspection in ranks or standby	**.45**	-.04	.40
25. Participated in CTC/NTC/JRTC rotation or FTX over 30 days	**.43**	-.21	.12
28. Prepared a lesson plan	**.43**	.07	.26
44. Acted as assistant commander at funeral detail or other public ceremony	**.42**	-.03	-.06
21. Total time spent in duty position one grade higher than actual grade	**.41**	.12	.10
41. Led/commanded Soldiers in drill and ceremony activities	**.39**	.01	.23
46. Appeared before a Soldier of the Month (or equivalent) Board	**.38**	.08	.14
32. Been part of a crew to perform Table VIII, Table XII, or TCPC	**.36**	-.14	-.02
43. Performed as Color Guard	**.36**	.05	-.07
27. Deployed on peace-keeping mission	**.34**	-.08	.03
26. Deployed on combat mission	**.32**	.00	.03
24. Total time in a unit specialty assignment	**.28**	.22	-.04
17. Served as a member of a unit advisory council or committee	**.16**	.16	.14
7. Used Windows Office programs to do job tasks	.05	**.78**	.06
1. Used a PC, Mac, or laptop	-.01	**.76**	-.01
3. Used the Internet for job or training requirements	.03	**.73**	.06
2. Communicated using e-mail	-.02	**.72**	-.02
4. Used the Windows NT operating system	.02	**.63**	.03
6. Troubleshooted a computer system malfunction	.00	**.62**	-.00
5. Operated an Army-specific computer system	-.01	**.36**	-.01
8. Trained or assigned as an I/O on any computer based simulator	.15	**.15**	.02
11. Established goals or other incentives to motivate subordinates	-.05	-.01	**.86**
12. Corrected unacceptable conduct of a subordinate	-.05	-.06	**.83**
10. Provided performance feedback to subordinates	-.01	.02	**.79**
14. Conducted formal inspection of subordinates' completed work	.03	-.04	**.77**
13. Trained other Soldiers in a task or procedure	.00	.02	**.74**
16. Counseled subordinates with disciplinary problems	.03	-.04	**.69**
15. Counseled subordinates regarding career planning	.11	.04	**.67**
9. Assigned to duty position with a responsibility for supervising 2+ Soldiers	.16	-.08	**.65**
18. Applied and supervised all 8 steps of troop leading procedures	.28	-.03	**.52**
40. Prepared a written counseling statement	.41	-.03	**.43**
20. Requested additional training opportunities	-.02	.18	**.35**
29. Led a PT class	.26	-.05	**.34**
19. Volunteered for additional duties/assignments	-.02	.18	**.31**

Note. $n = 1{,}882$.

ExAct Scoring

Based on the results of these EFAs, we formed three ExAct scores for subsequent validation: (a) Computer Experience (formed by averaging the standardized values from items 1 through 8), (b) Supervisory Experience (formed by averaging the standardized values from items 9 through 16, 18, 29, 40, and 42), and (c) General Experience (formed by averaging the standardized values from items 17, 19 through 28, 30 through 39, 41 and 43 through 46).[16] Items underlying the General Experience score reflected a variety of experiences that Soldiers tend to accumulate as they progress through their Army career (e.g., received and implemented a written operations order).

To evaluate these scores, coefficients alpha, item-deleted coefficients alpha, and score intercorrelations were computed. Table 5.2 presents the alpha coefficients and intercorrelations for the ExAct scores broken down by pay grade. All alphas indicated good internal consistency (Nunnally, 1978). Furthermore, item-deleted alphas indicated that removing items would not result in significant improvements in internal consistency (e.g., maximum observed increment = .03). Content analysis of the items suggested they were conceptually consistent with their respective composites. Therefore, all items were retained and scored. The moderate intercorrelations among ExAct scores offer evidence for the discriminant validity of ExAct scores and lend further support to the three-factor solution (D. Campbell & Fiske, 1959).

Table 5.2. ExAct Score Intercorrelations and Reliability Estimates

Predictor	ExAct Comp Exp $n_i = 8$	ExAct Sup Exp $n_i = 12$	ExAct Gen Exp $n_i = 26$
E4 Soldiers			
ExAct Computer Experience	(.84)		
ExAct Supervisory Experience	.06	(.89)	
ExAct General Experience	.20*	.66*	(.85)
E5 Soldiers			
ExAct Computer Experience	(.82)		
ExAct Supervisory Experience	.06*	(.84)	
ExAct General Experience	.19*	.48*	(.82)
E6 Soldiers			
ExAct Computer Experience	(.77)		
ExAct Supervisory Experience	.08*	(.82)	
ExAct General Experience	.24*	.41*	(.80)

Note. $n_{E4} = 444$; $n_{E5} = 880$; $n_{E6} = 556$. "n_i" indicates the number of items for each ExAct score. Correlations are uncorrected. Internal consistency reliability estimates (coefficients alpha) are in parentheses.

[16] Although items 19 and 20 loaded much higher on the Supervisory Experience factor than on the General Experience factor, they were included as part of the General Experience score. We hypothesized that their loading on the Supervisory Experience factor may be more reflective of their grouping with supervisory items on the ExAct form (i.e., order effects) rather than of their content similarity.

*$p < .05$ (one-tailed).

Descriptive Statistics

Descriptive statistics for the three ExAct scores, presented by subgroup (pay grade, race, gender, and CMF cluster) are presented in Tables 5.3 through 5.8. Raw and conditional effect sizes were calculated using the methods described in Chapter 3.

Table 5.3. Subgroup Differences by Pay Grade, Gender, and Race for ExAct Computer Experience

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	76	0.11	0.61	0.55	<.001	66	-0.02	0.62	0.13	.470
Male	362	-0.29	0.73			317	-0.12	0.69		
Race										
Black	92	-0.21	0.7	0.03	.786	89	-0.12	0.64	-0.14	.324
White	296	-0.24	0.73			294	-0.02	0.69		
E5										
Gender										
Female	109	0.24	0.49	0.47	<.001	90	0.16	0.48	0.12	.515
Male	767	-0.07	0.67			673	0.09	0.63		
Race										
Black	242	0.01	0.65	0.08	.279	241	0.08	0.59	-0.13	.294
White	523	-0.04	0.64			522	0.16	0.62		
E6										
Gender										
Female	57	0.36	0.43	0.29	.037	46	0.38	0.45	0.11	.612
Male	498	0.20	0.58			431	0.32	0.57		
Race										
Black	183	0.22	0.59	0.09	.330	182	0.36	0.58	0.04	.808
White	298	0.16	0.57			295	0.34	0.55		
Grade										
E5	880	-0.03	0.65	0.25	<.001	763	0.12	0.61	0.28	.007
E4	444	-0.21	0.73			383	-0.07	0.68		
E6	556	0.22	0.57	0.38	<.001	477	0.35	0.56	0.37	<.001
E5	880	-0.03	0.65			763	0.12	0.61		
E6	556	0.22	0.57	0.58	<.001	477	0.35	0.56	0.62	<.001
E4	444	-0.21	0.73			383	-0.07	0.68		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Given the number of effect sizes presented in Tables 5.3 through 5.9, only a few notable findings are summarized here. As expected, there were sizable differences in means for ExAct Supervisory Experience and ExAct General Experience across pay grades. Such findings support the validity of these scores as measures of Soldiers' military experience. Surprisingly, larger gender differences in ExAct Supervisory Experience and ExAct General Experience scores were generally found after controlling for race, CMF cluster, and pay grade differences. Specifically, women tended to score 0.56 (E4 Soldiers) to 0.73 (E6 Soldiers) standard deviation lower than men on ExAct Supervisory Experience, and 0.50 (E4 Soldiers) to 1.27 (E6 Soldiers) standard deviations lower than men on ExAct General Experience (holding race and CMF cluster constant).

Validity Estimates

Evidence for criterion-related validity was examined by computing zero-order correlations between the ExAct scores and the four criterion scores described in Chapter 3. Separate correlations were computed for E5 and E6 Soldiers, and differences between corresponding correlations (across pay grades) were tested for statistical significance. All correlations were corrected for criterion unreliability and direct range restriction on the predictor. Raw and corrected correlations are presented in Table 5.9.

The Computer Experience score was significantly predictive of observed performance for E5 Soldiers (r = .14) but not E6 Soldiers (r = .10), and exhibited low but statistically significant levels of validity for predicting expected future performance for E5 (r = .14) and E6 (r = .21) Soldiers. No significant E5/E6 differences were observed between corresponding correlations involving the Computer Experience score.

The Supervisory Experience score exhibited moderate, statistically significant levels of validity for predicting both observed and expected future performance for E5 Soldiers (.21 for observed performance, .30 for expected performance) but little validity for E6 Soldiers (-.03 for observed performance, .05 for expected performance). Although these differences between E5 and E6 correlations appear sizable, they were not statistically significant. A similar pattern of validity estimates was obtained for predicting the single-item criteria (Senior NCO Potential Rating and Overall Effectiveness Rating). The observed differences between E5 and E6 correlations may stem from a range restriction problem. For example, variation in the level of supervisory experience for E6 Soldiers may be less meaningful because most staff sergeants will have relatively high levels of supervisory experience. Sergeants, on the other hand, may vary more across the full spectrum of supervisory experience, and such variation (i.e., variation extending to lower levels of experience) may be particularly useful for predicting performance.

The General Experience score exhibited a pattern of validity similar to that of the Supervisory Experience score. For example, the General Experience score showed moderate, statistically significant validity estimates for predicting both observed and expected future performance for E5 Soldiers (.19 for observed performance, .20 for expected performance), but lower validity for E6 Soldiers (.10 and .11, respectively). Nevertheless, differences between the E5 and E6 correlations were not statistically significant. Again, the observed differences between E5 and E6 correlations may stem from a lesser degree of meaningful variation in General Experience among staff sergeants compared to sergeants.

Table 5.4. Differences between CMF Clusters for ExAct Computer Experience

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	69	56	0.33	0.30	0.61	0.64	—	-0.26	-0.95 **	-0.78 **	.	-0.94 **
2. INT	21	20	0.13	0.12	0.57	0.58	-0.27	—	-0.69	-0.52	.	-0.69
3. CBO	174	158	-0.41	-0.35	0.69	0.68	-1.01 **	-0.74 **	—	0.17	.	0.01
4. LOG	141	118	-0.29	-0.23	0.73	0.73	-0.84 **	-0.57 **	0.17	—	.	-0.17
5. CPA	12	10	0.06	0.08	0.50	0.56	.	.	.	.	—	.
6. COM	24	21	-0.30	-0.34	0.73	0.68	-0.85 **	-0.59 *	0.16	-0.01	.	—
Overall	444		-0.21		0.73							
E5 Soldiers												
1. ADM	84	70	0.47	0.46	0.51	0.48	—	-0.56	-0.97 *	-0.71 **	-0.48 *	-0.59 **
2. INT	37	35	0.13	0.12	0.59	0.60	-0.51 **	—	-0.42	-0.16	0.07	-0.04
3. CBO	330	285	-0.24	-0.13	0.68	0.68	-1.09 **	-0.57 **	—	0.26	0.49	0.38
4. LOG	289	248	-0.03	0.03	0.61	0.58	-0.75 **	-0.24	0.33 **	—	0.23	0.12
5. CPA	79	73	0.11	0.17	0.55	0.56	-0.55 **	-0.04	0.53 **	0.20	—	-0.11
6. COM	59	52	0.14	0.10	0.58	0.59	-0.50 **	0.01	0.58 **	0.25	0.05	—
Overall	880		-0.03		0.65							
E6 Soldiers												
1. ADM	60	51	0.46	0.43	0.42	0.45	—	.	-0.42	-0.44 *	-0.10	-0.30
2. INT	21	16	0.64	0.65	0.45	0.49	0.32	—	.	.	.	.
3. CBO	210	183	0.12	0.20	0.64	0.65	-0.59 **	-0.91 **	—	-0.02	0.32	0.12
4. LOG	169	146	0.13	0.18	0.55	0.53	-0.57 **	-0.89 **	0.02	—	0.35	0.14
5. CPA	65	55	0.31	0.38	0.39	0.41	-0.26	-0.58 **	0.33 *	0.31 *	—	-0.21
6. COM	31	26	0.32	0.26	0.48	0.52	-0.25	-0.57 *	0.34	0.32	0.01	—
Overall	556		0.22		0.57							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 5.5. Subgroup Differences by Pay Grade, Gender, and Race for ExAct Supervisory Experience

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	76	-1.26	0.75	-0.49	<.001	66	-1.41	0.79	-0.56	<.001
Male	362	-0.87	0.79			317	-0.98	0.76		
Race										
Black	92	-0.92	0.78	0.04	.720	89	-1.14	0.76	0.16	.128
White	296	-0.95	0.81			294	-1.26	0.77		
E5										
Gender										
Female	109	0.17	0.52	-0.23	.032	90	-0.04	0.53	-0.57	.013
Male	767	0.27	0.43			673	0.20	0.41		
Race										
Black	242	0.28	0.45	0.08	.296	241	0.12	0.44	0.18	.229
White	523	0.24	0.44			522	0.04	0.42		
E6										
Gender										
Female	57	0.22	0.48	-0.40	.006	46	0.06	0.49	-0.73	.014
Male	498	0.37	0.36			431	0.31	0.35		
Race										
Black	183	0.37	0.37	0.08	.404	182	0.22	0.36	0.21	.274
White	298	0.34	0.38			295	0.15	0.37		
Grade										
E5	880	0.25	0.44	1.50	<.001	763	0.08	0.43	1.67	<.001
E4	444	-0.95	0.80			383	-1.20	0.76		
E6	556	0.35	0.38	0.23	<.001	477	0.18	0.37	0.25	<.001
E5	880	0.25	0.44			763	0.08	0.43		
E6	556	0.35	0.38	1.63	<.001	477	0.18	0.37	1.81	.062
E4	444	-0.95	0.80			383	-1.20	0.76		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 5.6. Differences between CMF Clusters for ExAct Supervisory Experience

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	69	56	-1.15	-1.12	0.74	0.78	—	-0.16	-0.23	0.04	.	-0.36
2. INT	21	20	-1.31	-1.24	0.61	0.63	-0.20	—	-0.07	0.20	.	-0.21
3. CBO	174	158	-0.68	-1.30	0.83	0.80	0.58 **	0.78 **	—	0.27	.	-0.14
4. LOG	141	118	-1.03	-1.09	0.73	0.73	0.15	0.35	-0.43 **	—	.	-0.40 *
5. CPA	12	10	-0.93	-1.04	0.73	0.64	.	.	.	.	—	.
6. COM	24	21	-1.39	-1.40	0.73	0.74	-0.30	-0.10	-0.88 **	-0.45 *	.	—
Overall	444		-0.95		0.80							
E5 Soldiers												
1. ADM	84	70	0.11	0.08	0.56	0.57	—	-0.05	-0.64	0.37 *	0.21	0.06
2. INT	37	35	-0.03	0.06	0.57	0.48	-0.32	—	-0.58	0.42	0.26	0.11
3. CBO	330	285	0.37	-0.19	0.34	0.34	0.58 **	0.90 **	—	1.00 *	0.85	0.70
4. LOG	289	248	0.25	0.24	0.45	0.46	0.32 *	0.64 **	-0.27 **	—	-0.15	-0.31
5. CPA	79	73	0.18	0.17	0.51	0.50	0.16	0.48	-0.42 **	-0.16	—	-0.15
6. COM	59	52	0.09	0.11	0.35	0.37	-0.06	0.26	-0.64 **	-0.38 **	-0.22	—
Overall	880		0.25		0.44							
E6 Soldiers												
1. ADM	60	51	0.19	0.14	0.42	0.42	—	.	-0.71	0.44	0.57	0.18
2. INT	21	16	0.22	0.23	0.36	0.38	0.09	—	.	.	.	.
3. CBO	210	183	0.45	-0.12	0.31	0.30	0.68 **	0.59 **	—	1.15 *	1.28 *	0.89
4. LOG	169	146	0.32	0.30	0.41	0.41	0.36 *	0.27	-0.33 **	—	0.12	-0.26
5. CPA	65	55	0.39	0.35	0.35	0.37	0.54 **	0.45	-0.14	0.19	—	-0.38
6. COM	31	26	0.22	0.21	0.36	0.39	0.08	-0.02	-0.61 **	-0.28	-0.47 *	—
Overall	556		0.35		0.38							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 5.7. Subgroup Differences by Pay Grade, Gender, and Race for ExAct General Experience

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	76	-0.75	0.38	-0.44	<.001	66	-0.80	0.35	-0.50	.006
Male	362	-0.56	0.43			317	-0.58	0.42		
Race										
Black	92	-0.68	0.39	-0.25	.032	89	-0.73	0.38	-0.22	.127
White	296	-0.57	0.43			294	-0.64	0.42		
E5										
Gender										
Female	109	-0.14	0.38	-0.56	<.001	90	-0.27	0.37	-0.82	<.001
Male	767	0.09	0.41			673	0.05	0.39		
Race										
Black	242	0.03	0.39	-0.13	.091	241	-0.13	0.37	-0.11	.354
White	523	0.08	0.41			522	-0.09	0.39		
E6										
Gender										
Female	57	0.02	0.41	-1.18	<.001	46	0.00	0.37	-1.27	<.001
Male	498	0.41	0.33			431	0.40	0.31		
Race										
Black	183	0.33	0.40	-0.20	.043	182	0.18	0.34	-0.12	.491
White	298	0.40	0.34			295	0.22	0.30		
Grade										
E5	880	0.06	0.41	1.52	<.001	763	-0.11	0.39	1.41	<.001
E4	444	-0.59	0.43			383	-0.69	0.41		
E6	556	0.37	0.36	0.75	<.001	477	0.20	0.31	0.80	<.001
E5	880	0.06	0.41			763	-0.11	0.39		
E6	556	0.37	0.36	2.25	<.001	477	0.20	0.31	2.16	<.001
E4	444	-0.59	0.43			383	-0.69	0.41		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 5.8. Differences between CMF Clusters for ExAct General Experience

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	69	56	-0.74	-0.74	0.33	0.33	—	-0.07	-0.32	0.14	.	-0.08
2. INT	21	20	-0.71	-0.77	0.33	0.34	0.09	—	-0.25	0.21	.	-0.01
3. CBO	174	158	-0.46	-0.88	0.44	0.43	0.66 **	0.57 **	—	0.46	.	0.24
4. LOG	141	118	-0.65	-0.69	0.42	0.43	0.21	0.13	-0.44 **	—	.	-0.22
5. CPA	12	10	-0.22	-0.28	0.52	0.47	.	.	.	.	—	.
6. COM	24	21	-0.73	-0.78	0.35	0.37	0.04	-0.05	-0.61 **	-0.17	.	—
Overall	444		-0.59		0.43							
E5 Soldiers												
1. ADM	84	70	-0.06	-0.09	0.42	0.40	—	-0.28	-0.44	0.09	0.69 **	-0.35
2. INT	37	35	-0.12	-0.20	0.39	0.41	-0.13	—	-0.16	0.37	0.98 **	-0.07
3. CBO	330	285	0.17	-0.26	0.39	0.38	0.57 **	0.70 **	—	0.53	1.14 **	0.09
4. LOG	289	248	-0.01	-0.06	0.41	0.39	0.12	0.25	-0.45 **	—	0.61 **	-0.44 *
5. CPA	79	73	0.22	0.18	0.40	0.38	0.67 **	0.81 **	0.11	0.56 **	—	-1.05 **
6. COM	59	52	-0.12	-0.23	0.33	0.32	-0.14	-0.01	-0.71 **	-0.26	-0.82 **	—
Overall	880		0.06		0.41							
E6 Soldiers												
1. ADM	60	51	0.14	0.11	0.32	0.25	—	.	-0.15	0.03	1.01 **	0.61
2. INT	21	16	0.27	0.16	0.37	0.39	0.36	—	.	.	.	.
3. CBO	210	183	0.52	0.07	0.29	0.28	1.05 **	0.69 **	—	0.18	1.15 *	0.76
4. LOG	169	146	0.23	0.12	0.37	0.37	0.27 *	-0.10	-0.79 **	—	0.97 **	0.58 *
5. CPA	65	55	0.50	0.43	0.31	0.31	1.00 **	0.64 *	-0.05	0.74 **	—	-0.40
6. COM	31	26	0.41	0.30	0.30	0.27	0.75 **	0.38	-0.31	0.48 *	-0.26	—
Overall	556		0.37		0.36							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 5.9. Corrected and Raw Correlations between ExAct Scores and Criteria for E5 and E6 Soldiers

Criterion	Predictor		
	ExAct Computer Experience	ExAct Supervisory Experience	ExAct General Experience
E5 Soldiers			
Observed Performance Composite	.14 (.09*)	.21 (.08*)	.19 (.13*)
Expected Future Performance Composite	.14 (.08*)	.30 (.11*)	.20 (.12*)
Senior NCO Potential Rating	.07 (.05)	.19 (.07*)	.10 (.06)
Overall Effectiveness Rating	.06 (.04)	.21 (.08*)	.22 (.14*)
E6 Soldiers			
Observed Performance Composite	.10 (.07)	-.03 (-.02)	.10 (.07)
Expected Future Performance Composite	.21 (.12*)	.05 (.03)	.11 (.06)
Senior NCO Potential Rating	.03 (.02)	-.05 (-.03)	.01 (.01)
Overall Effectiveness Rating	.08 (.05)	-.04 (-.03)	.05 (.03)

Note. n_{E5} = 605-610; n_{E6} = 393-399. Correlations corrected for criterion unreliability and direct range restriction on the predictor appear outside of the parentheses. Raw correlations appear inside parentheses.

* $p < .05$ (one-tailed).

Differential Prediction Analyses

Table 5.10 presents the results of differential prediction analyses for ExAct scores by pay grade and criterion, examining gender and race as the demographic variables of interest. [17]

Overall, the results provide little evidence of differential prediction (i.e., slope bias). In the two cases where differential prediction was evident, the better prediction appeared to be for the minority group: the Supervisory Experience score was more predictive of observed performance for female E6 Soldiers (b = 0.15) than for male E6 Soldiers (b = -0.05), and the Computer Experience score was more predictive of expected future performance for black E6 Soldiers (b = 0.28) than for white E6 Soldiers (b = 0.07).

Intercept bias emerged only for gender-based comparisons when predicting expected future performance. Specifically, women had expected future performance composite scores that were roughly 0.33 to 0.38 point lower than men (at mean levels of the ExAct scores). These findings suggest that the ExAct Experience scores would tend to *overpredict* females' expected future performance if a common regression equation were used.

[17] All ExAct scores were standardized within pay grade to ease interpretation of the unstandardized regression weights prior to conducting these analyses. The demographic variables were coded as follows for purposes of analysis: race (white = 0, black = 1), gender (male = 0, female = 1).

Table 5.10. Differential Prediction Analyses for ExAct Scores

Criterion/Predictor	Demographic Main Effect		ExAct Score Main Effect				*r*			
			Gender		Race		Gender		Race	
	Gender	Race	M	F	W	B	M	F	W	B
			Observed Performance Composite							
E5 Soldiers										
ExAct Computer Experience	-.17	.01	.10	.00	.08	.12	**.12**	.00	**.09**	**.14**
ExAct Supervisory Experience	-.16	-.01	.09	-.01	.06	.09	**.10**	-.02	.06	.11
ExAct General Experience	-.12	.01	.11	.07	.09	.14	**.13**	.08	**.11**	**.17**
E6 Soldiers										
ExAct Computer Experience	-.09	-.13	.07	-.10	.04	.15	**.09**	-.10	.05	**.19**
ExAct Supervisory Experience	-.06	-.12	-.05	$.15_a$	-.01	.07	-.07	.24	-.01	.08
ExAct General Experience	-.01	-.11	.03	.11	.09	.09	.04	.15	**.12**	.12
			Expected Future Performance Composite							
E5 Soldiers										
ExAct Computer Experience	-.33*	.01	.12	-.12	.08	.07	**.12**	-.09	.08	.07
ExAct Supervisory Experience	-.37*	-.01	.13	-.04	.11	.10	**.13**	-.05	**.11**	.11
ExAct General Experience	-.37*	.02	.11	-.01	.12	.12	**.12**	-.01	**.12**	.13
E6 Soldiers										
ExAct Computer Experience	-.38*	-.15	.14	-.14	.07	$.28_a$	**.16**	-.09	.08	**.27**
ExAct Supervisory Experience	-.38*	-.14	-.01	.10	.04	.13	-.01	.10	.04	.12
ExAct General Experience	-.36	-.12	.02	.05	.10	.08	.02	.05	.11	.08

Note. Regression analysis sample sizes: $n_{E5\ Gender}$ = 603-608; $n_{E5\ Race}$ = 522-525; $n_{E6\ Gender}$ = 392-398; $n_{E6\ Race}$ = 343-348. Smaller sample sizes underlie the reported correlations because they were calculated for each subgroup separately. The "a" subscripts on the ExAct main effect values indicate the ExAct-by-demographic interaction term was statistically significant, $p < .05$ (two-tailed). Subscripts are located on the subgroup with the higher value. Correlations are uncorrected. Bolded correlations are statistically significant, $p < .05$ (one-tailed).

*$p < .05$ (two-tailed).

Summary

The ExAct scores showed more promise as predictors for future E4-to-E5 NCO promotion decisions than for future E5-to-E6 promotion decisions. Validity estimates tended to be higher for E5 Soldiers than for E6 Soldiers, particularly for the Supervisory and General Experience scores. The Computer Experience score yielded low (but statistically significant) validity estimates across pay grades.

Subgroup analyses revealed that women had significantly lower ExAct Supervisory and General Experience scores than men. These differences were sizable even after controlling for other demographic variables (i.e., race, pay grade, CMF cluster).

Overall, the ExAct scores were not differentially predictive for comparisons based on gender and race. However, there was evidence of intercept bias for gender (females' performance being overpredicted) when expected future performance was the criterion.

CHAPTER 6: SITUATIONAL JUDGMENT TEST

Gordon W. Waugh
HumRRO

Overview

Situational judgment tests assess the effectiveness of examinees' judgments about the appropriate courses of action in various job-related scenarios. Two such tests were developed for the NCO21 project. The Situational Judgment Test (SJT) comprises items measuring the eight NCO21 KSAs below.

- Directing, Monitoring, and Supervising Individual Subordinates
- Training Others
- Team Leadership
- Concern for Soldiers' Quality of Life
- Cultural Tolerance
- Motivating, Leading, and Supporting Individual Subordinates
- Relating to and Supporting Peers
- Problem-Solving/Decision Making Skill

These KSAs were selected based on the extent to which (a) they were identified as measurable by the SJT and (b) the SJT would, in combination with other measures, provide adequate coverage of the KSAs identified as critical in Phase II of the NCO21 research program.

A second test, the SJT-X, comprises items measuring *Knowledge of Inter-Relatedness of Units*. The SJT-X is separate from the SJT for two reasons: (a) its development process differed from the SJT, and (b) the items in the SJT-X contain lengthy scenarios—some requiring two pages of text. In contrast, SJT scenarios are typically about three sentences long.

Situational Judgment Test (SJT)

Instrument Description

The SJT form used in the concurrent validation had 40 items. Each item presented a 2–4 sentence scenario (i.e., description of a problem situation) followed by four possible actions (see Figure 6.1). Soldiers were instructed to indicate (a) which action was *most* effective and (b) which action was *least* effective. Each of the eight KSAs was represented by five items. The development of the SJT is described in Knapp et al. (2002).

When the final SJT scores were computed, only 24 items were included in the total score. Thus, the total SJT score for each Soldier was based on a shortened 24-item form. Two separate 24-item forms were used: one form for E4 and E5 Soldiers and a different form for E6 Soldiers. Twelve items appeared on both forms. Each of the eight KSAs was represented by three items. Because of the low construct validity and reliability of the eight scale scores, only total scores were used in the analyses.

Scoring

This section briefly describes how the SJT is scored. The development of the scoring process is described in detail later. The SJT scoring key is based upon SME ratings of the effectiveness of each response option. These ratings were obtained from 72 sergeants major (i.e., E9s), 3 E8 Soldiers, and 13 E7 Soldiers. Each SME rated only *some* of the options. Therefore, the number of SME ratings per option varies.

The score for an item is computed by subtracting the keyed effectiveness (i.e., the SMEs' mean effectiveness rating) of the option selected by the Soldier as *least* effective from the keyed effectiveness of the option selected as *most* effective. The total score for the test is the mean of the item scores.

One of your fellow Soldiers feels like he doesn't have to pitch in and do the work that you were all told to do. What should you do?

a. Explain to the Soldier that he is part of a team and needs to pull his weight.

b. Report him to the NCO in charge.

c. Find out why the Soldier feels he doesn't need to pitch in.

d. Keep out of it; this is something for the NCO in charge to notice and correct.

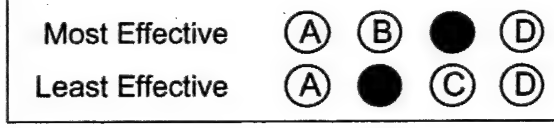


Figure 6.1. Example of a completed SJT item.

Comparison of Field Test Form and Validation Form

The SJT forms used in the field test vs. the validation differed in four major ways:

- Field test: two overlapping forms of 44 items each; validation: one 40-item form.
- Field test: 4–7 response options per item; validation: all items have 4 options.
- Field test: Soldiers rated the effectiveness of each action *and* picked the best and worst actions; validation: Soldiers just picked the best and worst actions.
- Field test: Soldiers wrote their responses (rating values and option letters) in the SJT item booklet; validation: Soldiers filled in circles on a scannable answer sheet.

Results

Data Preparation

SJT data were collected from 1,891 Soldiers. Before conducting analyses on the SJT dataset, two types of data cleaning were performed. First, 40 Soldiers were excluded from the SJT analyses for various reasons. A Soldier was dropped if he/she picked the same response option

(i.e., the same option letter) for more than 20 consecutive items. Five Soldiers exhibited such responding. Thirty-one Soldiers were dropped because they had more than four missing responses. In addition, 15 Soldiers who did not finish the test (i.e., did not complete the last item) were dropped because retaining them might have distorted the statistics for the last few items on the test. Some Soldiers were flagged by more than one of the exclusion rules.

Second, missing values were imputed. Because Soldiers with more than four missing responses were dropped, no more than four item scores were imputed for any Soldier. Item scores were imputed using regression (see Chapter 2).[18] A total of 360 values were imputed (0.49% of the item scores). Because of the extremely small percentage of imputed scores, the imputation process was unlikely to distort the results of the SJT analyses. After these steps, 1,851 Soldiers remained in the SJT database.

Score Development

After many analyses, the test characteristics and scoring algorithm shown below were adopted. The rationale behind these decisions is elaborated in the following sections.

- The Project A scoring algorithm (item's score equals key value for option picked as best minus key value for option picked as worst)
- Two test forms: one for E4 and E5 Soldiers, one for E6 Soldiers
- 24 items on each form
- 3 items per KSA
- Total SJT scores only reported (no scale scores)

Selection of the scoring algorithm. The field test results indicated that the best scoring algorithm was the one used in Project A (keyed value for option picked as best minus keyed value for option picked as worst; J. Campbell & Knapp, 2001). Two algorithms assessed in the field test could not be computed for the validation because of the different response format. Six other algorithms were added for the validation. The algorithms compared for the validation were as follows:

1. One point for identifying the best response (item score can be 0 or 1).
2. One point for identifying the worst response (item score can be 0 or 1).
3. Sum of algorithms 1 and 2 (item score can be 0, 1, or 2).
4. Minus one point for identifying the keyed worst response as the best (score was then reversed by multiplying it by –1 so that the item score can be 0 or 1).
5. Minus one point for identifying the keyed best response as the worst (score was then reversed by multiplying it by –1 so that the item score can be 0 or 1).
6. Sum of algorithms 4 and 5 (item score can be 0, 1, or 2).
7. Sum of algorithms 1, 2, 4, and 5 (item score can be 0, 1, 2, 3, or 4).

[18] Imputation was performed only for the final scoring algorithm.

8. Keyed effectiveness for the response picked as best (item score ranges from 1 to 7).
9. Keyed effectiveness for the response picked as worst (item score ranges from 1 to 7). This score was reversed by subtracting it from 8 so that higher scores are better.
10. Keyed effectiveness for response picked as best minus keyed effectiveness of response picked as worst (item score ranges from –6 to 6).

Table 6.1 shows the correlations among these 10 scoring algorithms. All but a few correlations are high, and some are very high. Thus, it appears that the algorithms are measuring very similar things. There does appear to be a difference, however, between the algorithms that give points for identifying the best response (1, 4, and 8) and those that give points for identifying the worst response (2, 5, and 9). This correlation pattern implies that the ability to make good decisions (in terms of deciding what to do in a situation) is slightly different from the ability to avoid bad decisions. An exploratory factor analysis confirmed this. When these six scores were factor analyzed, a two-factor solution emerged. The two factors were correlated .62.

Table 6.1. Correlations among SJT Scoring Algorithms

Algorithm	1	2	3	4	5	6	7	8	9	10
1. Best										
2. Worst	.50									
3. Best + Worst	.82	.90								
4. 1 – Reverse Best (picked keyed worst as best)	.70	.52	.69							
5. 1 – Reverse Worst (picked keyed best as worst)	.55	.79	.79	.51						
6. 2 – (Reverse Best + Reverse Worst)	.71	.77	.86	.84	.89					
7. Best + Worst – Reverse Best – Reverse Worst	.81	.89	.99	.76	.85	.93				
8. Key Value of Best	.90	.55	.81	.85	.58	.81	.83			
9. 8 – Key Value of Worst	.51	.94	.87	.52	.89	.83	.89	.57		
10. Key Value of Best – Key Value of Worst	.76	.87	.95	.74	.85	.92	.97	.85	.92	

Note. Scores are based on all 40 items, n = 1,850. Missing item scores were imputed using the Soldier's mean item score. All correlations are significant at $p < .0001$.

Table 6.2 shows the internal consistency reliability estimates and the criterion-related validity estimates of the scoring algorithms. The reliability estimates exhibit two trends. First, the algorithms related to identifying the *worst* response had higher reliability estimates than the algorithms related to identifying the *best* response. Second, reliability increased as the amount of information used by the algorithm increased. For example, algorithm 1 (alpha = .56) identifies only whether the Soldier correctly identified the best response; algorithm 8 (alpha = .74), however, weights the response by its keyed effectiveness value. In addition, algorithms that are merely combinations of other scores have higher reliability estimates than any of their constituent scores.

The differences between the algorithms' validity estimates are small and, in most cases, not statistically significant. These similar validity estimates show that the superior reliability of some algorithms does not necessarily translate into higher validity. For example, algorithm 1 had the lowest reliability, but its validity is higher (although not significantly higher) than many of the other algorithms.

Algorithm 10 has the highest reliability (tied with algorithm 9) and validity estimates. On a rational basis, it appears to include more information than the other algorithms. It measures both the ability to pick the best action and the ability to avoid the worst action, plus it weights the score by the keyed effectiveness value. It is the only scale to include all three of these pieces of information. Therefore, algorithm 10 was used in all subsequent SJT analyses. Note, however, that algorithm 7 does almost as well as algorithm 10. This is not surprising considering that these two algorithms correlate .97. There is one potential advantage of algorithm 7: Rather than using the actual values of the SMEs' effectiveness ratings, it uses only the identities of the best and worst responses options. Thus, this scoring key would generalize better to other sets of SMEs.

Table 6.2. Validity and Internal Consistency Reliability of SJT Scoring Algorithms

Algorithm	Reliability (coefficient alpha)	Correlation with Observed Performance	Correlation with Future Performance
1. Best	.56	.17	.12
2. Worst	.75	.14	.11
3. Best + Worst	.78	.18	.14
4. 1 – Reverse Best (picked keyed worst as best)	.63	.11	.08
5. 1 – Reverse Worst (picked keyed best as worst)	.71	.14	.11
6. 2 – (Reverse Best + Reverse Worst)	.75	.15	.12
7. Best + Worst – Reverse Best – Reverse Worst	.81	.18	.14
8. Key Value of Best	.74	.17	.14
9. 8 – Key Value of Worst	.84	.15	.13
10. Key Value of Best – Key Value of Worst	.84	.19	.15

Note. Validity estimates are uncorrected. Scores are based on all 40 items. n = 1,567–1,658 for the reliability estimates. n = 981 for observed performance, n = 991 for future performance. All correlations are significant at $p < .01$.

Item selection. Each SJT item (whether selected from a previous project or written for this project) was placed (on a rational basis) into one of the eight KSAs the instrument was designed to measure. When the field test (Knapp et al., 2002) data were factor analyzed, this eight-KSA structure was not supported. That is, the items did not fit into their pre-assigned KSAs. We decided, however, to draw an equal number of items from each KSA when constructing the SJT form for the validation. This balanced approach would help to ensure that the test covers a broad range of content.

Analyses were performed to determine whether the SJT could be shortened from its original length of 40 items without drastically reducing its quality. Initially, we shortened the test

to 32 items using Method 4 below. The resulting test had 4 items per scale. We found that reliability suffered little by dropping from 5 items per scale to 3 items per scale. Reliability dropped considerably, however, when only 2 items per scale were retained. Eventually, we decided to consider whether other methods of shortening the test could improve the psychometric characteristics of the test.

There is no consensus among test developers about the best method for shortening a test. Therefore, we examined various methods for shortening the test. Stanton, Sinar, Balzer, and Smith (2002) evaluated several criteria for shortening a popular job satisfaction measure (the Job Descriptive Index). These item-reduction criteria fall into three categories: descriptive statistics (e.g., drop items with low variance), internal consistency (e.g., drop items with low item-total correlations), and relationships with external variables (e.g., drop items that do not correlate with measures of related constructs). We developed five hypothetical shortened forms. They were based on the following criteria for dropping items:

1. Drop items having the lowest correlations with the supervisory performance ratings (i.e., the lowest estimated validities).
2. Drop items having the lowest correlations with the other predictors.
3. Drop items that have the lowest combination of reliability, validity, and correlations with other predictors.
4. Drop items based on the item-scale and item-total correlations. Drop the item in the scale with the lowest item-*scale* correlation. If the two lowest values are similar, then drop the item among these two with the lowest item-*total* correlation. Repeat this process until the desired number of items remain in each scale.
5. Drop items with the lowest item-total correlations.

The best 24 items for each method were selected[19]. This test-length was chosen because the item-criterion correlations (i.e., Method 1) became rather small after the 24th item. A double cross-validation design was used to minimize capitalization on chance. The sample from the validation was randomly split into two equal samples. Sample 1 was used to select the items; sample 2 was used to compute the SJT validity estimates based on the selected items. Then the roles of samples 1 and 2 were reversed: Sample 2 was used to select the items; sample 1 was used to compute the SJT validities based on the selected items. Thus, each sample acted as both an analytic sample and a validation sample. The results are shown in Table 6.3.

These results suggest that selecting items based on their criterion-related validity leads to the highest cross-validated validity. Picking items based solely on their correlations with the total score (Method 5) yielded the lowest validity (although not significantly lower than most of the other methods). As mentioned, the items for the final E5 and E6 forms were selected based upon the item validities. When constructing the final forms, all E5 and all E6 Soldiers were used to compute the estimated item validities. That is, the E5 and E6 groups were split into two random samples *only* for the cross-validation analysis.

[19] Method 4, however, used 32 items. This was initially considered the final form before the other methods were considered. Thus, Method 4 operates a baseline for evaluating the other methods.

Table 6.3. Criterion-Related Validity Estimates of Different Item-Selection Methods

Item Selection Method	Mean Validity in the 2 Analytic Samples	Mean Validity in the 2 Validation Samples
1. Item-criterion correlations	.240	.197
2. Correlations with other predictors	.178	.182
3. Combination of methods 1, 2, & 5	.205	.181
4. Iterative removal of item with lowest item-scale correlation.	.167	.167 *
5. Item-total correlations	.150	.150 *

Note. The observed performance composite was used as the criterion. Methods are listed in descending order of validity in the validation sample. Methods 1, 2, 3, and 5 are based on a 24-item test. Method 4 is based upon 32 items; its validity would likely be lower if based upon a 24-item test. $n = 485$ (sample 1) and 486 (sample 2). Correlations are uncorrected. Item selection using Method 4 was done using samples 1 and 2 combined (i.e., it was not cross-validated).

* Cross-validated estimate is significantly different from the Method 1 cross-validated estimate at $p < .05$.

To select the items for each of the two 24-item forms, the two items within each scale with the lowest item-criterion correlations were dropped. Additional item analyses were performed solely to help decide which items to select. For these analyses, a single criterion was required. For the purpose of these analyses, the observed and future performance ratings were given equal weight. Thus, the criterion score was the average of the observed and future performance rating composites. These analyses were conducted separately for E5 and E6 Soldiers. Thus, separate shortened test forms were created for E5 Soldiers and for E6 Soldiers. The two 24-item test forms had 12 items in common. Scores for E4 Soldiers were computed using the E5 test form.

Reliability Estimates

Scales. The internal consistency reliability of the test was estimated using coefficient alpha (see Table 6.4). The low reliability estimates for the scales is not surprising considering that each contains only three items. Because of these low reliability estimates, only the total SJT score was used in the SJT analyses.

Total score. The reliability estimates for the total scores are not very high, but they are typical for situational judgment tests. Even at the item level, situational judgment tests are multidimensional and heterogeneous by nature. That is, a typical item measures more than one construct and the items measure the various constructs to different degrees. Internal consistency reliability estimates, on the other hand, assume that a single construct or the same set of constructs (to the same degree) underlies the items. Thus, coefficient alpha usually underestimates the reliability of situational judgment tests. Test-retest estimates of reliability are preferred, but they could not be obtained in this validation. Considering these limitations, the reliability estimates for the total SJT scores in Table 6.4 are respectable. They are high enough to show that a common set of constructs underlies most of the items.

Table 6.4. Internal Consistency Reliability Estimates for the SJT

	Scale	24-item form			40-item form		
		E4	E5	E6	E4	E5	E6
	Sample Size	437	866	545	437	866	545
1	Relating to and Supporting Peers	.46	.44	.40	.52	.56	.51
2	Cultural Tolerance	.47	.32	.05	.55	.41	.26
3	Motivating, Leading, and Supporting Individual Subordinates	.29	.20	.24	.39	.36	.35
4	Training Others	.17	.19	.30	.42	.25	.35
5	Directing, Monitoring, and Supervising Individual Subordinates	.17	.19	.05	.33	.28	.32
6	Concern for Soldiers' Quality of Life	.22	.27	.27	.37	.40	.38
7	Problem-Solving/Decision Making Skill	.10	.13	.12	.26	.18	.24
8	Team Leadership	.38	.34	.33	.41	.38	.42
	Total Score	.76	.73	.68	.85	.82	.81

Note. For the 24-item test, E4 and E5 Soldiers were scored using the E5 form, whereas E6 Soldiers were scored using the E6 form. All Soldiers completed the *same* 40-item form.

Dimensionality

Scale intercorrelations for the 24-item forms are shown in Tables 6.5 and 6.6. The correlations among the scales are relatively low. These low correlations are likely due to the low scale reliabilities. For example, when corrected for unreliability, the correlation of .34 between Peers and Cultural Tolerance in Table 6.5 becomes .81. Because the à priori scales are not being measured reliably, the scale scores cannot form the underlying dimensions of the item scores.

A factor analysis had previously been performed on the 40-item test for all Soldiers combined. Principal axis factor extraction was selected. To help determine the number of factors to extract, a parallel analysis was performed using Monte Carlo methods. That is, factor analysis was conducted on 100 random datasets, each with the same sample size and same number of variables as the target dataset. The scree plot of each random dataset was compared to that from the actual dataset. The factor number just before the scree plots crossed was noted for each pair. In most pairs, the two scree plots crossed between the 22nd and 23rd factors, thus, suggesting a 22-factor solution. Because the SJT was intended to measure eight constructs, an eight-factor solution was run using oblique rotation (which did not restrict the size of the factor intercorrelations). The solution was uninterpretable. In sum, no meaningful factor solution could be obtained.

Descriptive Statistics

The means and standard deviations for the SJT total score are shown in Table 6.7. Each Soldier's total score was computed two ways: once using the E5 form and once using the E6 form. Table 6.7 shows that the E6 form was slightly more difficult than the E5 form (dependent *t*-tests found the difference to be statistically significant at $p < .0001$ for each pay grade).

Table 6.5. Correlations Among the SJT Scales: E4 and E5 Soldiers

	Scale	Peers	Cult	Motiv	Train	Super	QLife	DM	Lead
1	Relating to and Supporting Peers								
2	Cultural Tolerance	.34							
3	Motivating, Leading, and Supporting Individual Subordinates	.36	.35						
4	Training Others	.26	.19	.19					
5	Directing, Monitoring, and Supervising Individual Subordinates	.37	.31	.35	.21				
6	Concern for Soldiers' Quality of Life	.39	.28	.29	.24	.30			
7	Problem-Solving/Decision Making Skill	.24	.20	.20	.16	.21	.21		
8	Team Leadership	.47	.37	.37	.23	.33	.33	.24	
	Total Score	.72	.62	.62	.49	.62	.61	.53	.67

Note. n = 1,303. The 24-item E5 form was used for both the E4 and E5 Soldiers. All correlations are significant at $p < .0001$.

Table 6.6. Correlations Among the SJT Scales: E6 Soldiers

	Scale	Peers	Cult	Motiv	Train	Super	QLife	DM	Lead
1	Relating to and Supporting Peers								
2	Cultural Tolerance	.25							
3	Motivating, Leading, and Supporting Individual Subordinates	.33	.24						
4	Training Others	.17	.08	.23					
5	Directing, Monitoring, and Supervising Individual Subordinates	.25	.17	.23	.16				
6	Concern for Soldiers' Quality of Life	.33	.27	.31	.30	.19			
7	Problem-Solving/Decision Making Skill	.20	.16	.18	.11	.09	.16		
8	Team Leadership	.39	.14	.32	.22	.17	.24	.20	
	Total Score	.65	.51	.65	.50	.48	.58	.52	.58

Note. n = 545. The 24-item E6 form was used. Correlations greater than .08 are significant at $p < .05$; correlations greater than .11 are significant at $p < .01$.

Table 6.7. Descriptive Statistics by Pay Grade for the Total Score of the SJT

		E5 Form		E6 Form	
Pay Grade	n	M	SD	M	SD
E4	437	1.80	0.70	1.70	0.68
E5	866	2.19	0.57	2.06	0.55
E6	545	2.38	0.50	2.24	0.48

Table 6.7 shows the means and standard deviations of the total SJT scores computed by gender, pay grade, and race. Because the cell sizes were unbalanced, conditional means (i.e., estimated least square means) were computed. Conditional means prevent unbalanced cell sizes from causing misleading results (see Appendix C). Unless otherwise noted, the discussion of group differences refers to conditional means rather than raw means.

Tables 6.8 and 6.9 show the group differences by gender, race, and pay grade for the E5 and E6 forms, respectively. Females and males did not differ significantly. Among E4 Soldiers, the size of the advantage for females was meaningful but not statistically significant. Among E4 and E5 Soldiers, whites significantly outperformed blacks by 0.32 and 0.29 standard deviation, respectively. The advantage for E6 whites (using the E6 form) was 0.35 standard deviation, but this difference was nonsignificant. These differences are small compared to tests of general cognitive ability, in which whites usually outperform blacks by about 1.0 standard deviation.

Differences by pay grade were also computed. The higher pay grades scored significantly higher than the lower grades. One would expect a Soldier's standing on the constructs targeted by the SJT to improve with training and experience (especially in supervision and leadership). Therefore, this result provides evidence of the construct validity of the SJT.

The mean difference between the E4 and E5 levels was double that between the E5 and E6 levels. This is what one would expect because the amount of training and the number of experiences related to leadership increase much more from E4 to E5 than from E5 to E6. The promotion from E4 to E5 involves profound change—from Soldier to NCO. In contrast, promotion from E5 to E6 increases the NCO's span of control and brings some new experiences, but the types of tasks performed are similar.

There is also significant leadership training when a Soldier moves to the E5 level. To be promoted to E5, one must have attended PLDC (although currently that can be waived for 1 year following the promotion date). Normally, Soldiers would attend PLDC as a very senior E4 Soldier or right after becoming an E5 Soldier. Although it is only a 30-day course, PLDC is a total immersion experience. Much of the instruction is academic, but the Soldiers constantly rotate through different levels of leadership assignments during the period, changing positions and responsibilities daily. One day a Soldier might be a section leader, the next day a Company Commander. For many Soldiers, this is their first time in a leadership position. Even for E4 Soldiers who have had temporary leadership roles, the PLDC experience is intense and has a lasting effect.

The SJT scores were also compared by CMF (see Tables 6.10 and 6.11). None of the differences among the conditional means is significant. Although some of the effect sizes are moderate, the small sample sizes might have prevented them from reaching statistical significance.

Table 6.8. Subgroup Differences by Pay Grade, Gender, and Race for the Total Score on the E5 Form of the SJT

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	76	2.03	0.63	0.40	0.002	66	1.98	0.67	0.31	0.071
Male	355	1.75	0.70			311	1.77	0.69		
Race										
Black	89	1.66	0.73	–0.26	0.034	85	1.77	0.72	–0.32	0.017
White	294	1.84	0.68			292	1.99	0.68		
E5										
Gender										
Female	108	2.28	0.49	0.17	0.099	89	2.22	0.51	–0.01	0.976
Male	754	2.18	0.57			663	2.22	0.55		
Race										
Black	239	2.12	0.58	–0.20	0.014	238	2.14	0.56	–0.29	0.029
White	515	2.23	0.54			514	2.30	0.54		
E6										
Gender										
Female	57	2.44	0.38	0.14	0.29	46	2.39	0.37	0.04	0.866
Male	487	2.37	0.51			421	2.37	0.50		
Race										
Black	177	2.27	0.60	–0.39	<.001	176	2.29	0.60	–0.45	0.017
White	294	2.44	0.42			291	2.47	0.42		
Grade										
E5	866	2.19	0.57	0.57	<.001	752	2.22	0.54	0.50	<.001
E4	437	1.80	0.70			377	1.88	0.69		
E6	545	2.38	0.50	0.33	<.001	467	2.38	0.49	0.30	0.010
E5	866	2.19	0.57			752	2.22	0.54		
E6	545	2.38	0.50	0.83	<.001	467	2.38	0.49	0.73	<.001
E4	437	1.80	0.70			377	1.88	0.69		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 6.9. Subgroup Differences by Pay Grade, Gender, and Race for the Total Score on the E6 Form of the SJT

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	76	1.92	0.61	0.38	.002	66	1.88	0.66	0.29	.086
Male	355	1.65	0.69			311	1.68	0.67		
Race										
Black	89	1.57	0.70	–0.26	.033	85	1.67	0.70	–0.33	.015
White	294	1.74	0.66			292	1.88	0.66		
E5										
Gender										
Female	108	2.13	0.47	0.14	.159	89	2.13	0.48	0.06	.753
Male	754	2.05	0.56			663	2.10	0.54		
Race										
Black	239	2.02	0.56	–0.14	.089	238	2.07	0.54	–0.19	.148
White	515	2.09	0.53			514	2.17	0.53		
E6										
Gender										
Female	57	2.31	0.42	0.16	.257	46	2.29	0.43	0.12	.608
Male	487	2.23	0.49			421	2.24	0.48		
Race										
Black	177	2.15	0.58	–0.35	.002	176	2.19	0.58	–0.35	.061
White	294	2.30	0.41			291	2.34	0.41		
Grade										
E5	866	2.06	0.55	0.54	<.001	752	2.12	0.53	0.51	<.001
E4	437	1.70	0.68			377	1.78	0.67		
E6	545	2.24	0.48	0.33	<.001	467	2.27	0.48	0.28	.014
E5	866	2.06	0.55			752	2.12	0.53		
E6	545	2.24	0.48	0.80	<.001	467	2.27	0.48	0.73	<.001
E4	437	1.70	0.68			377	1.78	0.67		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 6.10. Differences between CMF Clusters for the Total Score on the E5 Form of the SJT

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	68	55	1.89	1.85	0.68	0.71	—	0.36	–0.40	0.04	.	.
2. INT	21	20	2.08	2.10	0.60	0.45	0.28	—	–0.75	–0.31	.	.
3. CBO	170	155	1.70	1.58	0.75	0.75	–0.28 *	–0.56 **	—	0.44	.	.
4. LOG	140	118	1.81	1.88	0.68	0.65	–0.12	–0.40 *	0.16	—	.	.
5. CPA	12	10	2.02	1.91	0.53	0.48	.	.	.	.	—	.
6. COM	23	19	1.91	1.94	0.52	0.50	0.02	–0.26	0.30	0.14	.	—
Overall	437		1.80		0.70							
E5 Soldiers												
1. ADM	82	68	2.30	2.33	0.52	0.48	—	–0.03	–0.76	–0.24	–0.15	–0.08
2. INT	37	35	2.38	2.32	0.33	0.33	0.14	—	–0.73	–0.21	–0.12	–0.05
3. CBO	323	279	2.10	1.92	0.61	0.57	–0.36 **	–0.51 **	—	0.52	0.61	0.68
4. LOG	284	245	2.20	2.20	0.57	0.55	–0.19	–0.33 **	0.18 *	—	0.09	0.16
5. CPA	79	73	2.27	2.25	0.56	0.58	–0.07	–0.21	0.30 *	0.12	—	0.07
6. COM	59	52	2.34	2.29	0.41	0.43	0.06	–0.08	0.43 **	0.25	0.13	—
Overall	866		2.19		0.57							
E6 Soldiers												
1. ADM	59	50	2.43	2.45	0.42	0.42	—	.	–0.55	–0.05	–0.14	0.10
2. INT	21	16	2.42	2.37	0.37	0.39	–0.02	—	.	.	.	.
3. CBO	205	179	2.32	2.18	0.55	0.55	–0.23	–0.21	—	0.50	0.41	0.65
4. LOG	165	142	2.39	2.42	0.50	0.48	–0.09	–0.07	0.14	—	–0.10	0.15
5. CPA	65	55	2.40	2.38	0.47	0.43	–0.08	–0.05	0.15	0.01	—	0.25
6. COM	30	25	2.50	2.50	0.48	0.44	0.14	0.16	0.37	0.23	0.21	—
Overall	545		2.38		0.50							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 6.11. Differences between CMF Clusters for the Total Score on the E6 Form of the SJT

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	68	55	1.82	1.80	0.69	0.72	—	0.22	–0.42	–0.03	.	.
2. INT	21	20	1.91	1.95	0.60	0.44	0.14	—	–0.65	–0.25	.	.
3. CBO	170	155	1.59	1.52	0.72	0.72	–0.33 *	–0.46 *	—	0.39	.	.
4. LOG	140	118	1.72	1.78	0.65	0.62	–0.15	–0.28	0.18	—	.	.
5. CPA	12	10	1.93	1.83	0.46	0.43	.	.	.	.	—	.
6. COM	23	19	1.77	1.78	0.63	0.64	–0.07	–0.20	0.26	0.08	.	—
Overall	437		1.70		0.68							
E5 Soldiers												
1. ADM	82	68	2.16	2.20	0.52	0.49	—	0.01	–0.55	–0.25	–0.10	–0.07
2. INT	37	35	2.24	2.21	0.39	0.41	0.14	—	–0.56	–0.26	–0.11	–0.08
3. CBO	323	279	1.98	1.91	0.59	0.56	–0.32 **	–0.47 **	—	0.30	0.45	0.48
4. LOG	284	245	2.06	2.07	0.54	0.53	–0.18	–0.32 **	0.15	—	0.15	0.18
5. CPA	79	73	2.14	2.15	0.54	0.56	–0.04	–0.18	0.29 *	0.14	—	0.03
6. COM	59	52	2.17	2.16	0.46	0.50	0.02	–0.13	0.34 *	0.19	0.05	—
Overall	866		2.06		0.55							
E6 Soldiers												
1. ADM	59	50	2.30	2.31	0.48	0.49	—	.	–0.31	–0.09	–0.08	–0.03
2. INT	21	16	2.31	2.29	0.37	0.40	0.02	—	.	.	.	.
3. CBO	205	179	2.21	2.16	0.50	0.50	–0.19	–0.22	—	0.22	0.23	0.28
4. LOG	165	142	2.25	2.27	0.46	0.45	–0.12	–0.14	0.08	—	0.01	0.05
5. CPA	65	55	2.26	2.27	0.50	0.47	–0.10	–0.12	0.10	0.02	—	0.04
6. COM	30	25	2.27	2.29	0.50	0.49	–0.08	–0.10	0.12	0.04	0.02	—
Overall	545		2.24		0.48							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Validity Estimates

Criterion-related validity was computed using separate forms for E5 and E6 Soldiers (see Table 6.12). These correlations were corrected for criterion unreliability and range restriction. The SJT's correlations with observed performance ratings were .39 and .25 for E5 and E6 Soldiers, respectively. As explained earlier, the 24 items for the final E5 and E6 scores were selected according to their correlations with the two criteria (observed performance composite and expected future performance composite). Therefore, the reported validity estimates are somewhat inflated because the same sample was used to select the items and compute the validities of the total scores. Based on cross-validation results, the shrunken validities are estimated to be .32 and .17 for the E5 and E6 forms, respectively.

Table 6.12. Corrected and Raw Correlations between the SJT and Criteria for E5 and E6 Soldiers

Criterion	Not Corrected for Shrinkage	Corrected for Estimated Shrinkage
E5 Soldiers		
Observed Performance Composite	.39 (.23)	.32
Expected Future Performance Composite	.37 (.19)	.29
Senior NCO Potential Rating	.28 (.16)	
Overall Effectiveness Rating	.36 (.19)	
E6 Soldiers		
Observed Performance Composite	.25 (.16)	.17
Expected Future Performance Composite	.28 (.16)	.19
Senior NCO Potential Rating	.18 (.10)	
Overall Effectiveness Rating	.16 (.10)	

Note. n_{E5} = 595–600; n_{E6} = 386–391. All correlations are significant at $p < .05$ (one-tailed). Correlations corrected for direct range restriction on the predictor and criterion unreliability appear outside of the parentheses. Raw correlations appear inside parentheses.

Differential Prediction Analyses

Fairness analyses were conducted to determine whether the SJT-criterion prediction equation differed across gender or race. Results of these analyses are shown in Table 6.13. Multiple moderated regression (MMR), based on the Cleary (1968) model of fairness, was used to compute the results (see Chapter 4). Table 6.13 presents the results of differential prediction analyses for SJT scores by pay grade and criterion, examining gender and race as the demographic variables of interest.[20]

[20] To ease interpretation of the unstandardized regression weights, all SJT scores were standardized within pay grade prior to conducting these MMR analyses. The demographic variables were coded as follows for purposes of analysis: race (white = 0, black = 1); gender (male = 0, female = 1).

Values under the *r* column represent the within-group correlations between the SJT scores and the performance ratings. Correlations can be interpreted as the amount of increase in one variable (in SD units) reflected by a 1.0 standard deviation increase in the other variable. The variables are standardized within each group.

There were no significant main effects for race. For gender, three of the four main effects were significant. The future performance ratings showed the largest differences. At the same SJT score, females' unstandardized future performance ratings were 0.43 and 0.44 point below the males' ratings for E5 and E6 Soldiers, respectively. Thus, females' SJT scores actually *overpredicted* their future performance. For the observed performance ratings, the difference (which reflected overprediction for females) was significant only for E5 Soldiers.

The SJT predicted performance significantly better for one group than for the other group in only one of eight comparisons. Specifically, among E5 Soldiers, the SJT predicted future performance better for whites ($b = 0.25$) than for blacks ($b = 0.07$).

Table 6.13. Differential Prediction Analyses for the SJT

	Demographic Main Effect		SJT Score Main Effect				*r*			
			Gender		Race		Gender		Race	
Criterion/Pay Grade	Gender	Race	M	F	W	B	M	F	W	B
Expected Future Performance Comp.										
E5 Soldiers	–.43*	.03	.18	.30	$.25_a$	.07	.23	.33	.24	.08
E6 Soldiers	–.44*	–.09	.16	.12	.12	.21	.17	.12	.12	.22
Observed Performance Comp.										
E5 Soldiers	–.23*	.04	.19	.34	.22	.17	.19	.27	.24	.21
E6 Soldiers	–.12	–.11	.13	.10	.09	.18	.18	.10	.11	.25

Note. $n_{E5\ Gender}$ = 593–598; $n_{E5\ Race}$ = 515–518; $n_{E6\ Gender}$ = 336–340; $n_{E6\ Race}$ = 385–390. The "a" subscripts on the SJT main effect values indicate that the SJT-by-demographic interaction term was statistically significant (which indicates that the two groups have different slopes). Subscripts are located on the subgroup with the higher value.

*$p < .05$ (two-tailed) for the demographic main effect.

Situational Judgment Test X (SJT-X)

Instrument Description

Type of Items

The purpose of the SJT-X is to measure *Knowledge of the Inter-Relatedness of Units*. This knowledge is believed to be much more important to performance at the NCO level in the future Army than in the current Army. The SJT-X comprises only three items but its scenarios (i.e., situation descriptions) average 700 words in length. Because only 24 Soldiers took the field test version of the SJT-X, no changes were made to the instrument before the validation.

Response Format

Reliability was a concern because there were so few items. To maximize reliability, we wanted to obtain as many responses as possible from the Soldiers. Thus, Soldiers completed the SJT-X by reading each scenario and rating the effectiveness of each action listed (i.e., response option) on a 7-point scale. This response format generated many more responses and scores (26 responses) than simply asking the Soldiers to pick the best and worst action for each item (6 responses). All else being equal, the greater number of responses should increase reliability. It also allowed us to compute scoring algorithms based on the Soldiers' effectiveness ratings. Soldiers also indicated the *most effective* response and the *least effective* response for each item.

Figure 6.2 shows an example of a completed SJT-X item. The example is intended to illustrate only the format of the SJT-X items. This example is much shorter than any of the SJT-X items. Development of the SJT-X is described in Knapp et al. (2002).

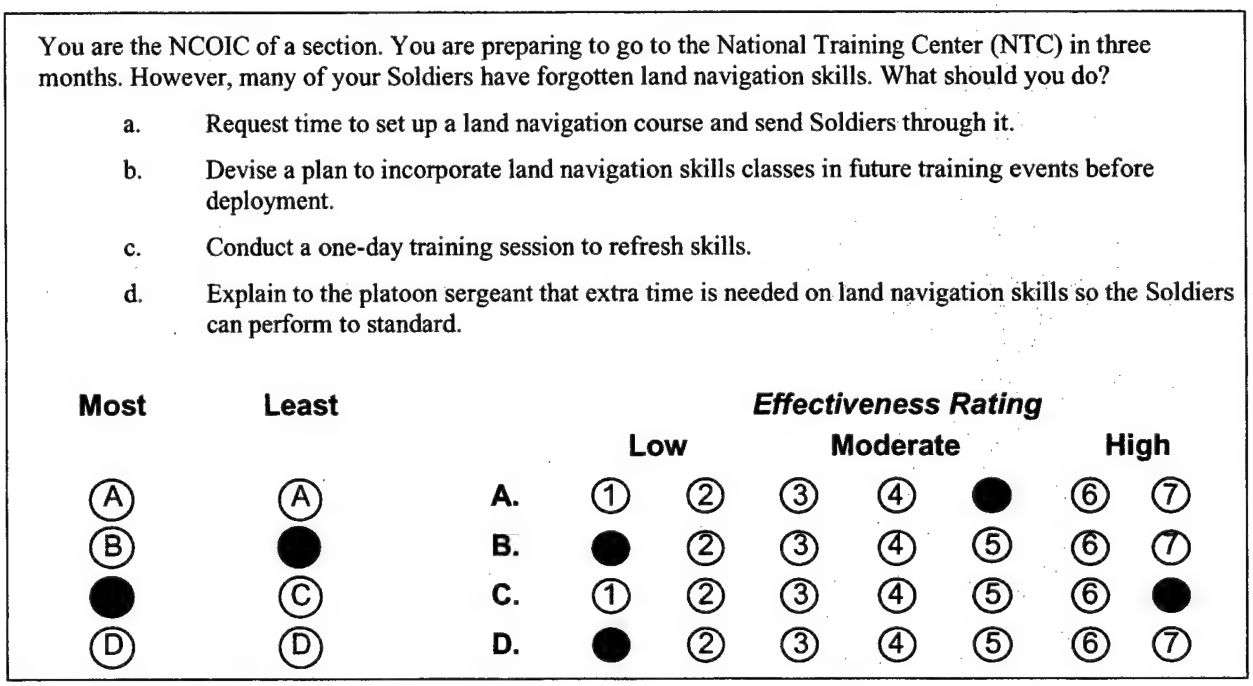


Figure 6.2. Format of SJT-X items.

Results

Data Preparation

SJT-X data were collected from 525 Soldiers. The SJT-X was administered only to E6 Soldiers because very few E5 Soldiers would have been exposed, either through experience or training, to the types of situations contained in the SJT-X. Before conducting analyses on the SJT-X dataset, 55 Soldiers were excluded from the SJT-X analyses because they did not complete the SJT-X or showed questionable response patterns. Specially, a Soldier was dropped if he or she gave the same effectiveness rating for more than 10 options in a row. Seventeen

Soldiers exhibited such responding. Soldiers were also dropped if they had more than one missing response within any of the three items. Forty-nine Soldiers had too many missing responses and were dropped from the analyses. The three Soldiers who did not finish the test (i.e., did not complete the last item) were dropped because retaining them might have distorted the statistics for the test. Some Soldiers were flagged by more than one of the exclusion rules.

After data cleaning, 64 missing values were imputed. Because Soldiers with more than three missing responses were dropped (i.e., more than one missing response per item), no more than three item scores were imputed for any Soldier. Imputation was performed only for each of the scoring algorithms. The score for an option was computed as the mean of other option scores within the item. A total of 64 values were imputed (0.52% of the option scores). Because of the extremely small percentage of imputed scores, the imputation process was unlikely to distort the results of the SJT-X analyses. After these procedures, 470 Soldiers remained in the SJT-X database.

Selecting the Scoring Algorithm

Three algorithms were used to compute the scores for the SJT-X. Algorithm 1 was examined for the SJT validation (see Table 6.1), and algorithm 2 was examined in the SJT field test (Knapp et al., 2002). The other algorithms used in the SJT research were not tried because they proved, in the SJT analyses, to be very similar to one of these two algorithms. The third algorithm was unique to the SJT-X validation. The algorithms were as follows:

1. Algorithm 10 from the SJT: The SMEs' effectiveness value of the action picked as best minus the SMEs' effectiveness value of the action picked as worst.

2. Algorithm 4 from the SJT field test: The absolute difference between the Soldier's effectiveness rating for the option and the official effectiveness value for the option (i.e., the SMEs' mean rating). The score for an item was simply the sum of the option scores.

3. Correctness of Option Rank-Ordering[21]: The absolute difference between the Soldier's ranking of an option and the SMEs' ranking of the option. Thus, the maximum score is achieved when the Soldier puts the options in the same order (in terms of effectiveness) as the SMEs. These absolute differences were summed to produce a total score for the item. The item score was rescaled so that a score of 0 represented random responding and a score of 1 represented a perfect score.

The third algorithm, which evaluated *correctness of option rank-ordering*, resulted in the highest criterion-related validity (see Table 6.14). This algorithm was computed only after the other two algorithms produced disappointing validity results. One problem with Algorithm 1 is that a Soldier's score suffers when he or she uses a (a) narrower or wider range of ratings than the SMEs or (b) different mean rating than the SMEs. That is, a Soldier can rank-order the

[21] Algorithm 3 could not be used with the SJT because the SJT form did not ask soldiers to rate each option. The SJT asked the soldier only to pick the best and worst (of four) options. Thus, it is not known how the soldier would have ranked the two unpicked options.

options perfectly (in terms of effectiveness) but get a low score on the item because, for example:

1. The Soldier's ratings range from 3–5 whereas the SMEs' ratings range from 1–7, or
2. The Soldier's ratings range from 1–4 whereas the SMEs' ratings range from 4–7.

One could argue that what is important in dealing with a situation is simply picking the best thing to do rather than accurately evaluating the relative effectivenesses of the alternative actions. In the test, this is reflected by the ability to rank-order the options correctly. Therefore, it is not necessary to accurately determine the effectiveness of each possible action on an interval scale. In algorithms 1 and 2, when the mean or variance of a Soldier's effectiveness ratings differ from the mean or variance of the SMEs' ratings, this difference is treated as error. Algorithm 3 ignores these differences and considers only the rank ordering of the options.

Table 6.14. Estimated Validities of the SJT-X Scoring Algorithms

Algorithm	Correlation with Ratings of:	
	Observed Performance	Expected Future Performance
1. Absolute difference between Soldier's and SMEs' option effectiveness ratings	.06	.05
2. Absolute difference in SME means between Soldier's picks of best and worst options	.11	.07
3. Difference between Soldier's and SMEs' ranking of the options	.14	.15

Note. n = 342–347. Correlations are uncorrected. Algorithm 3 was chosen as the final scoring algorithm for the SJT-X.

Selection of Response Options

Analyses were performed to determine whether the SJT-X could be improved by dropping some of the response options. An option was dropped if its exclusion increased the internal consistency reliability of its item.[22,23] Options were dropped using an iterative process. In the first iteration, the option whose exclusion increased the item coefficient alpha the most was dropped. In step two, the same procedure was repeated using the remaining options. The process stopped when coefficient alpha could no longer be increased by dropping options. For item 1, 6 of the original 7 options were retained; for item 2, 5 of the 7 original options were retained; and for item 3, 5 of the original 12 options were retained. Thus, across all items, 16 of the original 26 options were retained.

[22] The internal consistency reliability estimate of an item was computed as follows. Each option has a score (its rank among the item's options). The internal consistency reliability estimate of an *item* is the internal consistency reliability estimate (computed using Cronbach's alpha) of the item's set of option scores.

[23] Because of the heterogeneous nature of situational judgment tests, coefficient alpha is a lower-bound estimate of reliability for the SJT-X. More appropriate reliability study designs (e.g., test-retest or alternate forms), however, could not be used.

After performing several validity and reliability analyses, we decided to use the test characteristics and scoring algorithm shown below. The justifications for these choices are presented in the following sections.

- 3 items
- Number of response options: 6, 5, and 5 for items 1, 2, and 3, respectively
- A total SJT-X score only (no item scores)
- Scoring algorithm 3: a rank order correspondence algorithm (option's score is the difference between the rank order of the option provided by the Soldier vs. the SMEs)

Reliability

Table 6.15 shows the internal consistency reliability estimates of the scoring algorithms. Coefficient alpha was computed in two ways for algorithms 1 and 3: (a) using the option scores and (b) using the item scores. Algorithm 2 does not compute option scores. The nature of the iterative process of dropping options caused the number of options to differ across the three algorithms. That is, options were no longer dropped when the internal consistency reliability of an item was maximized. This stopping point differed across algorithms.

The reliability estimates for the total scores are not very high. As explained previously, situational judgment tests typically have low internal consistency reliability because they are multidimensional even within each item. Because the SJT-X has only three items, it is not likely to achieve high reliability. The estimated reliabilities (based on the option scores) of algorithms 1 and 3 are adequate, however, for a short test that is used in conjunction with other tests to make decisions. A test-retest reliability design (with an interval of a few weeks between tests) would provide a better estimate of the test's reliability.

Table 6.15. Internal Consistency Reliability Estimates of the SJT-X Scoring Algorithms

Algorithm	*n* of Options	Reliability Based on Option Scores	Reliability Based on Item Scores
1. Difference between Soldier's and SMEs' option effectiveness ratings	23	.56	.50
2. Difference in SME means between Soldier's picks of best and worst options	26	—	.24
3. Difference between Soldier's and SMEs' ranking of the options	16	.63	.25

Note. $n = 342–347$. Correlations are uncorrected. Algorithm 3 was chosen as the final scoring algorithm for the SJT-X.

Descriptive Statistics

The means and standard deviations for the SJT-X total score are shown in Table 6.16. As mentioned previously, the item and total scores were scaled so that a score of 0 represents random responding and a score of 1 represents a perfect score (i.e., the Soldier's ranking of the options matched the SMEs' ranking of the options). Item 1 was much more difficult than the other two items. In addition, its standard deviation was quadruple that of the other items.

Table 6.16. Descriptive Statistics for the SJT-X

Item	*M*	*SD*
1	0.22	0.40
2	0.83	0.10
3	0.84	0.10
Total Score	0.63	0.15

Note. $n = 470$.

Dimensionality

The three SJT-X items had low intercorrelations (see Table 6.17), but the values are typical of most tests. Item 1 correlates .95 with the total SJT-X score because its high standard deviation gives it a much higher weight than the other two items when the total score is computed. Thus, the item 1 score is almost equivalent to the total score. The final set of response options was factor analyzed to determine what constructs might underlie the data. A parallel analysis was performed to help determine the number of factors in the data. We computed the eigenvalues for the correlation matrix of the 16 response option scores. The diagonal of the correlation matrix was replaced with multiple-squared correlations before the eigenvalues were computed. The parallel analysis indicated a 13-factor solution. Three additional rules of thumb for determining the number of factors were used. Nine eigenvalues exceeded zero (another criterion for determining the number of factors). Six eigenvalues exceeded 1.0 when there were ones in the diagonal of the correlation matrix. Finally, there was a large discontinuity in the scree plot (i.e., a sudden large drop in the eigenvalues) after the 12th eigenvalue. Thus, the test appears to be heterogeneous. This was expected because of the multidimensional nature of situational judgment tests, in general, and the complexity of the scenarios.

Tables 6.18 and 6.19 show the differences between groups in their total SJT-X scores. Means were computed by gender, race, and CMF. Because the cell sizes were unbalanced, conditional means (i.e., estimated least square means) were computed. There were no significant subgroup differences based on the conditional means. Although Combat Operations (in Table 6.19) appeared to have the lowest means—and some moderate effect sizes—the results were not statistically significant.

Table 6.17. Correlations Among the SJT-X Items

Item	Item 1	Item 2	Item 3
1			
2	.18		
3	.23	.15	
Total Score (corrected)	.27	.20	.25
Total Score (uncorrected)	.95	.40	.44

Note. $n = 470$. All correlations are significant at $p < .01$. When computing an item's correlation with the *Total Score (corrected)*, the total score omitted the target item. In contrast, the *Total Score (uncorrected)* included the target item.

Table 6.18. Subgroup Differences by Gender and Race for the SJT-X

Group	Raw					Conditional				
	n	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
Gender										
Female	44	0.59	0.16	–0.27	.090	35	0.61	0.16	–0.15	.603
Male	425	0.63	0.15			368	0.63	0.16		
Race										
Black	148	0.62	0.16	–0.07	.505	147	0.61	0.16	–0.08	.744
White	259	0.63	0.15			256	0.62	0.15		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 6.19. Differences between CMF Clusters for the SJT-X

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
1. ADM	45	39	0.62	0.63	0.16	0.14	—	.	–0.50	–0.09	–0.07	0.11
2. INT	17	13	0.59	0.66	0.14	0.14	.	—	.	.	.	.
3. CBO	172	149	0.65	0.55	0.15	0.16	0.21	.	—	0.41	0.43	0.61
4. LOG	146	127	0.61	0.61	0.15	0.16	–0.06	.	–0.27*	—	0.02	0.20
5. CPA	59	49	0.61	0.62	0.15	0.16	–0.09	.	–0.31*	–0.03	—	0.18
6. COM	31	26	0.65	0.64	0.18	0.19	0.16	.	–0.06	0.22	0.25	—
Overall	470		0.63		0.15							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Validity Estimates

The criterion-related validity of the SJT-X was estimated using four criteria (see Table 6.20). These values were quite respectable considering that the SJT-X has only three items. Most of the validity of the test can be attributed to the first item (see Table 6.21). Item 2 had moderate validity for all criteria except expected future performance. Item 3 had essentially no validity. Item 3 was by far the longest item (its scenario was almost two pages long) and it was the most complex. Although it is possible that the amount of reading or the number of things to consider might have been too much for the Soldiers, items 2 and 3 differ little in terms of their mean score and standard deviation. Because almost all of the SJT-X's validity is due to item 1, it could be improved by replacing items 2 and 3 with items of the same quality as item 1. This would also shorten the test considerably because the lengthy item 3 would be replaced with a shorter item. Alternatively, the length of the test could be maintained by replacing item 3 with two or three items.

Table 6.20. Corrected and Raw Correlations between the SJT-X and Criteria

Criterion	*r*
Expected Future Performance Composite	.22 (.15)
Observed Performance Composite	.18 (.14)
Senior NCO Potential Rating	.18 (.13)
Overall Effectiveness Rating	.15 (.11)

Note. n = 341–346. All correlations are significant at $p < .05$ (one-tailed). Correlations corrected for indirect range restriction on the predictor and criterion unreliability appear outside of the parentheses. Raw correlations appear inside parentheses.

Table 6.21. Correlations between the SJT-X Items and Criteria

Criterion	Total Score	Item 1	Item 2	Item 3
Expected Future Performance Composite	.15*	.16*	.04	.03
Observed Performance Composite	.14*	.13*	.11*	.01
Senior NCO Potential Rating	.13*	.12*	.11*	.02
Overall Effectiveness Rating	.11*	.09*	.12*	.02

Note. n = 341–346. Correlations are uncorrected.

* $p < .05$ (one-tailed).

The construct validity[24] of the SJT-X is difficult to assess for several reasons. First, because none of the other predictors was designed to measure *Knowledge of the Inter-Relatedness of Units*, no measures can be used to assess the convergent validity of the SJT-X. One of the observed performance rating scales, however, assesses the closely related construct *Coordination of Multiple Units and Battlefield Functions*. Second, E6 Soldiers do not currently need to exhibit *Knowledge of the Inter-Relatedness of Units* and do not have to deal with the situations described in the SJT-X. They are expected, however, to have to deal with these situations in the future Army.

A few of the ExAct items should be related to the SJT-X. For example, one would expect Soldiers who had been deployed to combat or peacekeeping missions, or who had issued or implemented operations orders, to score higher on the SJT-X. Table 6.22 shows these correlations. It also shows that the SJT-X has a low correlation with general cognitive ability.

The SJT-X scores were somewhat correlated with the observed performance rating scale *Coordination of Multiple Units and Battlefield Functions*. This shows convergent validity. To show good discriminant validity, the SJT-X should show consistently lower correlations with the other observed performance scales. This, however, was not the case: the SJT-X's correlation with *Coordination of Multiple Units and Battlefield Functions* was no higher than its correlations with many of the other performance scales. This result is inconclusive, however, because of the high

[24] Construct validity of the other predictors is discussed in Chapter 9.

correlations among the observed performance scales. That is, one would expect any measure to correlate similarly with most of the observed performance scales.

Table 6.22. Correlations Between the SJT-X and the Observed Performance Rating Scales

Measure	Correlation with SJT-X	*n*
ASVAB - General Technical	.09*	453
Observed Performance Ratings: Coordination of Multiple Units and Battlefield Functions	.13*	341
ExAct: Computer Experience Scale	.01	470
ExAct: Supervisory Experience Scale	.05	470
ExAct: General Experience Scale	.12*	470
ExAct: Deployed on a Combat Mission	–.06	469
ExAct: Deployed on a Peacekeeping Mission	.01	468
ExAct: Issued or Implemented an Operations Order (2 items)	.21*	470
SJT	.16*	460

* $p < .05$.

Contrary to expectations, the SJT-X was unrelated to deployment. The SJT-X had a moderate correlation with the two ExAct items related to issuing or implementing operations orders.

Differential Predication Analyses

Fairness analyses were conducted to determine whether the SJT-X-criterion prediction equation differed across gender or race. Table 6.23 presents the results of differential prediction analyses for SJT-X scores by criterion, examining gender and race as the demographic variables of interest.[25]

Table 6.23. Differential Prediction Analyses for the SJT-X

	Demographic Main Effect		SJT-X Score Main Effect				*r*			
			Gender		Race		Gender		Race	
Criterion/Pay Grade	Gender	Race	M	F	W	B	M	F	W	B
Observed Performance Composite	-.17	-.14	.10	.09	.11	.08	.14	.11	.16	.10
Expected Future Performance Composite	-.46*	-.18	.12	.21	.13	.14	.13	.16	.14	.14

Note. n_{Gender} = 340-345; n_{Race} = 296-301.

*$p < .05$ (two-tailed) for the demographic main effect.

[25] To ease interpretation of the unstandardized regression weights, all scores were standardized within pay grade prior to conducting these analyses. The demographic variables were coded as follows for purposes of analysis: race (white = 0, black = 1); gender (male = 0, female = 1).

Values under the *r* column represent the within-group correlations between the SJT-X scores and the performance ratings. Correlations can be interpreted as the amount of increase in one variable (in SD units) reflected by a 1.0 standard deviation increase in the other variable. The variables are standardized within each group.

There was only one significant effect: the demographic main effect for gender. At the same SJT-X score, females' unstandardized future performance ratings were 0.46 point below the males' ratings. Thus, females' SJT-X scores actually *overpredicted* their future performance.

Summary

The SJT's high validity supports its use in helping decide whom to promote to the E5 and E6 levels. These validity estimates are based upon 24-item forms, which would require only about 40 minutes to administer. Separate forms should be developed for the E5 and E6 levels. In addition, the effects of race and gender are relatively small. Females and blacks scored almost as high as males and whites, respectively. The differential prediction analyses showed no fairness problems with the SJT.

Although the SJT was developed for promotion purposes, it could also serve as a valuable training tool. The SJT could provide realistic scenarios to E4 and E5 Soldiers, which they could use to hone their decision-making skill.

The SJT-X targets a relatively narrow construct: *Knowledge of the Inter-Relatedness of Units*. The SJT-X had respectable criterion-related validity in spite of imperfect criteria. Although most supervisors provided performance ratings for a construct similar to this one (*Coordination of Multiple Units and Battlefield Functions*), it is unlikely that many supervisors have actually observed their subordinates in situations relevant to this construct. Thus, it is possible that the validity estimates would have been higher had a better criterion measure been available.

It appears that the SJT-X could be improved markedly by writing more items like item 1 and by avoiding lengthy, complex items like item 3. Thus, it is likely that a short SJT-X with respectable construct and criterion-related validity can be developed. It might not be possible to complete an accurate validation of SJT-X items until some groups of E5 and E6 Soldiers are given leadership roles in the situations depicted in the SJT-X. Thus, the SJT-X should probably not be implemented until that time.

CHAPTER 7: SEMI-STRUCTURED INTERVIEW

Gordon W. Waugh and Christopher E. Sager
HumRRO

Overview

The Army uses a Board interview as part of its current semi-centralized promotion system, but this interview is not highly structured nor is it intended to cover KSAs identified in the NCO21 project. Therefore, it seemed reasonable to design a semi-structured interview as another experimental predictor measure. The NCO21 semi-structured interview uses a standard protocol for conducting the interview, selecting questions from a question bank, developing new questions, and evaluating interviewees in several target areas. Project staff trained senior NCO interviewers to conduct the structured interviews.

The NCO21 interview covers the nine target areas listed below.

- Adaptability
- Level of Effort and Initiative on the Job
- Level of Integrity and Discipline on the Job
- Relating to and Supporting Peers
- Leadership Skills/Potential[26]
- Self-Management and Self-Directed Learning Skill[27]
- MOS/Occupation-Specific Knowledge and Skill
- Military Presence
- Oral Communication Skill

Instrument Description

Interview Components

The development of the NCO21 semi-structured interview is described in Knapp et al. (2002). Basic components of the interview include (a) a question bank, (b) target area definitions, (c) anchored rating scales for each of the nine target areas, (d) instructions and worksheet for developing questions to supplement the question bank, and (e) a worksheet on which to record and consolidate ratings from two interviewers.

There are 50 questions in the validation version of the interview question bank, each of which taps one of six target areas (see Table 7.1). There are no questions pertaining to *Military Presence* or *Oral Communication*, as these KSAs are evaluated based on the Soldier's overall performance throughout the interview. There are also no questions for *MOS/Occupational Knowledge and Skill,* with the understanding that interviewers will develop questions in this area themselves.

---

[26] Combination of three NCO21 KSAs: (a) *Motivating, Leading, and Supporting Individual Subordinates;* (b) *Team Leadership*; and (c) *Directing, Monitoring, and Supervising Individual Subordinates.*
[27] Combination of two NCO21 KSAs: (a) *General Self-Management Skill* and (b) *Self-Directed Learning Skill.*

There are three types of interview questions: (a) past-experience questions, (b) hypothetical-situation questions, and (c) fact-based questions. Table 7.1 shows that nearly half (45%) of the question bank used in the validation data collection included past-experience questions, whereas 55% were hypothetical-situation questions. Fact-based questions are not suitable for the question bank because, in an operational setting, this easily would result in compromise. Therefore, the interviewers wrote all of the fact-based questions. In the earlier field test, hypothetical-situation questions were found to be more conducive to assessing interview performance in some categories (e.g., *Adaptability*, *Leadership Skills/Potential*) than others (e.g., *Self-Management/Self-Directed Learning*, *Relating to Peers*). The question bank contained a few questions not posed to E4 Soldiers because of the level of experience the questions presumed.

Table 7.1. Summary of Validation Data Collection Interview Scales and Questions

Scale	Total Number of Questions in Bank	Number of Past Experience Questions	Number of Hypothetical Situation Questions
1. Adaptability	9	2	7
2. Military Presence	N/A	—	—
3. Level of Effort & Initiative on the Job	4	2	2
4. Level of Integrity & Discipline on the Job	11	3	8
5. Relating to and Supporting Peers	7	5	2
6. Leadership Skills/ Potential	13	6	7
7. Oral Communication Skill	N/A	—	—
8. Self-Management/Self-Directed Learning Skill	6	5	1
9. MOS/Occupation-Specific Knowledge and Skill	Interviewer Writes	Interviewer Writes	Interviewer Writes
Total	50	23	27

Note. The interviewers assessed *Oral Communication Skills* and *Military Presence* by observing the Soldiers throughout the interview.

Interviewees are evaluated in the nine areas using structured rating scales. Each rating scale ranges from 1 (low effectiveness) to 7 (high effectiveness) and contains three anchor levels (i.e., low, moderate, and high). Each anchor includes (a) short descriptions about general behavior demonstrated at that level and (b) two to four specific behavioral examples of what the Soldier could have described in his or her response. The interview rating scales are very similar in format to the observed performance rating scales.

Other supporting materials developed for the interview include an interview script, suggestions for probing interviewees' responses, instructions for making ratings, and an interview worksheet to record ratings. The interview worksheet lists the nine areas covered in the interview, a place to record ratings (i.e., to circle a value from 1 to 7), and space to record notes.

Interview Process

For purposes of the validation research, the interview was designed for administration by pairs of senior NCOs. Procedures comparable to those described here could be developed for panels of more than two interviewers.

The most senior NCO in a pair, who served as the *lead interviewer*, was responsible for making introductions, explaining the process to the interviewee, and making the final decision on selecting interview questions. The second interviewer, designated the *recorder*, was responsible for consolidating the ratings at the end of the interview. Both interviewers could ask questions and were instructed to take notes during the interview. At the end of the interview, both interviewers reviewed their notes and made *independent* judgments using the target area rating scales (*pre-consensus* ratings). If their ratings differed by more than 2 points, then the interviewers discussed the interviewee's performance and revised their discrepant ratings to within 2 points (*post-consensus* ratings). The recorder averaged the two sets of post-consensus ratings to obtain an overall rating for each scale. An overall rating for the interview was computed by averaging the final scale ratings.

Each interview lasted approximately 20 minutes, with 10 additional minutes for completing the rating forms. After all site interviews had been conducted, the interviewers were asked to evaluate the interview and training by completing a rating form, answering open-ended questions, and writing comments.

Interviewer Training

A 3-hour training session and associated materials were developed to train senior NCOs to conduct the NCO21 interviews. The training consists of a lecture, observation and discussion of two mock interviews, and practice with feedback. A significant goal of the training is to distinguish the NCO21 interview from the board procedures with which the NCOs are accustomed, and to demonstrate the need for the interviewers to carefully adhere to the standardized procedures.

Validation Data Collection

Due to time and logistical limitations, the semi-structured interview was administered to E4 and E5 Soldiers only. Interview appointments were scheduled for some Soldiers, but most Soldiers were interrupted from their written session to do the interview.

One staff member of the project team served as Interview Manager. This individual led the interviewer training session and monitored the interview process throughout the course of the data collection period. The Interview Manager also designated the Soldiers to be interviewed by each interviewer pair, based on a match between the MOS of an interviewer and the Soldier when possible (thus allowing an interviewer to ask MOS-specific questions). Interviewers who were not in the same MOS as the interviewee did not pose MOS-specific questions.

There were 64 interviewers who formed 32 pairs. With one exception, each interviewer pair stayed together throughout the interviews. (At one site, two interviewers became a pair for just the last few interviewees when their partners left.) Each E4 or E5 Soldier was interviewed by

one interviewer pair. The number of Soldiers interviewed by each pair ranged from 14 to 43 Soldiers (excluding the pair that interviewed only a few Soldiers), with a mean of 30 and a standard deviation of 7. Across the seven sites, 45% of the interviewers were white, 87% were male, and 29 MOS were represented. The interviewers were in pay grades E7–E9: 77% E7, 19% E8, and 5% E9 (these sum to more than 100% because of rounding).

Results

There were three sets of scores on a 7-point scale for each E4 or E5 Soldier participating in the interview: one score for each of the nine scales from each of the two interviewers (pre-consensus ratings), a set of mean consensus ratings (post-consensus) for each scale, and overall interview score (i.e., mean of the mean consensus ratings).

Descriptive Statistics

Only two Soldiers were missing any interviewer ratings (excluding the MOS-specific knowledge scale). After dropping these Soldiers from the interview dataset, 944 Soldiers remained, of which 302 (32%) were E4 Soldiers and 641 (68%) were E5 Soldiers (the grade of one interviewee was unknown). Table 7.2 shows the mean consensus rating for each scale as well as the overall mean interview scores (i.e., composite scores). The composite interview score was computed as the mean of the consensus scale scores. A composite score excluding the MOS-specific rating was also computed, because (a) most Soldiers were not rated in this area and (b) Soldiers who were rated in this area were primarily evaluated by only one interviewer. Overall, the amount of variability in the ratings suggests interviewers were able to discriminate among Soldiers. The mean values (4.74–4.99) likely indicate some minor leniency in the ratings but the degree of leniency is lower than that found in most research on interview and performance ratings.

Table 7.2. Descriptive Statistics for the Semi-Structured Interview

Scale	*M*	*SD*
1. Adaptability	4.89	1.02
2. Military Presence	4.94	1.08
3. Level of Effort & Initiative on the Job	4.98	1.01
4. Level of Integrity & Discipline on the Job	4.88	1.21
5. Relating to and Supporting Peers	4.79	1.05
6. Leadership Skills/ Potential	4.84	1.11
7. Oral Communication Skill	4.99	1.06
8. Self-Management/Self-Directed Learning Skill	4.66	1.15
9. MOS/Occupation-Specific Knowledge and Skill (n = 296)	4.74	1.62
Composite Interview Score for Soldiers with a Rating for MOS/Occupation-Specific Knowledge and Skill (n = 296)	4.93	0.83
Composite Interview Score Excluding MOS-Specific Ratings	4.87	0.85

Note. n = 944 for all variables except where indicated. Interviewers' mean consensus ratings ranged from 1.0–7.0 for the scales.

To maximize sample size, the subgroup analyses used the composite interview score without the MOS-specific rating. Table 7.3 shows the subgroup mean differences by Soldiers' gender, race, and grade. The table has two sets of results. The left half of the table shows raw statistics (i.e., other variables were not controlled). The right half of the table shows conditional statistics. These represent the statistics obtained while controlling for the other variables in the subgroup analyses (see the description of conditional means in Appendix C). The two sets of statistics gave similar results for race and pay grade but differed for gender. The conditional means analyses revealed only one significant difference: As expected, E5 Soldiers received higher scores than did E4 Soldiers.

Table 7.3. Subgroup Differences by Pay Grade, Gender, and Race for the Semi-Structured Interview (Composite Score Excluding MOS-Specific Knowledge)

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	54	4.88	0.99	0.41	.009	44	4.80	0.99	0.18	.486
Male	243	4.53	0.84			211	4.65	0.85		
Race										
Black	59	4.58	0.92	0.01	.929	57	4.70	0.95	–0.06	.761
White	200	4.57	0.87			198	4.75	0.85		
E5										
Gender										
Female	90	4.95	0.84	–0.08	.498	73	4.79	0.84	–0.44	.085
Male	547	5.01	0.80			475	5.14	0.80		
Race										
Black	173	5.00	0.90	0.01	.939	173	4.94	0.89	–0.07	.702
White	376	4.99	0.77			375	4.99	0.76		
Grade										
E5	641	5.00	0.80	0.45	<.001	548	4.97	0.80	0.28	.044
E4	302	4.60	0.88			255	4.72	0.87		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 7.4 shows mean differences by CMF cluster. For E4 Soldiers, only three of the CMF categories had large enough samples to analyze (i.e., sample size of at least 20). There were no significant differences among the conditional means for these three CMF categories. For E5 Soldiers, five of the six CMF categories had samples large enough to analyze. There were no significant differences among the conditional means for these five CMF categories.

Table 7.4. Differences between CMF Clusters for the Semi-Structured Interview (Composite Score Excluding MOS-Specific Knowledge)

CMFC	*n*		*M*		*SD*		*Effect Size*					
	Raw	Con	Raw	Con	Raw	*Con*	*1. ADM*	*2. INT*	*3. CBO*	*4. LOG*	*5. CPA*	*6. COM*
E4 Soldiers												
1. ADM	44	33	4.77	4.81	1.14	1.26	—	.	–0.22	–0.12	.	.
2. INT	9	8	4.26	4.16	0.79	0.70	.	—	.	.	.	.
3. CBO	118	104	4.52	4.62	0.71	0.72	–0.28	.	—	0.10	.	.
4. LOG	99	84	4.60	4.70	0.90	0.87	–0.18	.	0.09	—	.	.
5. CPA	8	7	5.31	5.49	0.48	0.53	.	.	.	.	—	.
6. COM	22	19	4.53	4.56	1.10	1.04	–0.26	.	0.01	–0.08	.	—
Overall	302		4.60		0.88							
E5 Soldiers												
1. ADM	61	48	5.22	5.20	0.94	0.98	—	.	–0.48	–0.33	0.03	–0.20
2. INT	18	18	5.11	4.58	0.91	0.91	.	—	.	.	.	.
3. CBO	240	205	4.93	4.82	0.78	0.77	–0.36 *	.	—	0.15	0.50	0.28
4. LOG	237	200	4.95	4.94	0.80	0.81	–0.33 *	.	0.03	—	0.35	0.13
5. CPA	38	37	5.25	5.22	0.72	0.69	0.03	.	0.39 *	0.37 *	—	–0.23
6. COM	46	40	5.07	5.04	0.72	0.75	–0.18	.	0.18	0.15	–0.22	—
Overall	641		5.00		0.80							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race. All significance tests are two-tailed.

*$p < .05$. **$p < .01$.

Dimensionality

Mean consensus scores for each scale were correlated to assess the relationships among the scales. Table 7.5 shows that all scale intercorrelations were significant ($p < .0001$). The high correlations suggested that the semi-structured interview scales, except for *MOS/Occupation-Specific Knowledge and Skill*, measured a single construct or a set of highly-related constructs.

An exploratory factor analysis (EFA) using principal axis extraction was performed to determine if the semi-structured interview assessed more than one construct. A parallel analysis was run. First, a scree plot of the interview data was created by computing the eigenvalues of the reduced correlation matrix (i.e., with squared multiple correlations in the diagonal) of the eight interview scales (*MOS/Occupation-Specific Knowledge and Skill* was excluded). Second, 100 random datasets were created. Each dataset had eight variables and the same number of cases as the interview dataset. Third, the scree plot of the reduced correlation matrix was computed for each of these datasets. Finally, the scree plot of the actual data was compared with the scree plot of each of the 100 random datasets. In all cases, the scree plots crossed between the first and second factors. This indicates that one factor likely underlies the interview data. In addition, the correlation matrix of the interview data had only one eigenvalue greater than 1.00 (first two eigenvalues = 4.94, 0.63). These results, coupled with the high scale intercorrelations, strongly suggested that the semi-structured interview measures one underlying construct. We concluded that the overall composite score is the most appropriate summary score for the interview.

Table 7.5. Inter-Scale Correlations for the Semi-Structured Interview (Composite Score Excluding MOS-Specific Knowledge)

Scale	1	2	3	4	5	6	7	8	9
1. Adaptability	—								
2. Military Presence	.53	—							
3. Level of Effort & Initiative on the Job	.64	.57	—						
4. Level of Integrity & Discipline on the Job	.47	.48	.55	—					
5. Relating to and Supporting Peers	.59	.51	.62	.56	—				
6. Leadership Skills/ Potential	.61	.56	.65	.56	.67	—			
7. Oral Communication Skill	.60	.70	.60	.52	.59	.65	—		
8. Self-Management/Self-Directed Learning Skill	.54	.47	.54	.42	.46	.51	.54	—	
9. MOS/Occupation-Specific Knowledge and Skill	.37	.35	.38	.36	.41	.45	.49	.30	—

Note. $n = 944$ except for correlations with *MOS/Occupation-Specific Knowledge and Skill* ($n = 296$). All correlations are significant at $p < .0001$.

Reliability Estimates

Internal consistency reliability estimates (using Cronbach's alpha) were computed for two composite scores: one with and one without MOS-specific ratings. Computing the composite without the MOS-specific rating offered a greater sample size for the computation of internal consistency reliability estimates. Alpha was .89 ($n = 296$) for the composite score and .91 ($n = 944$) for the composite score excluding MOS-specific ratings. Based on this analysis, there was no evidence to suggest that any scales should be dropped from the semi-structured interview.

Interrater agreement across interviewer pairs was estimated using a generalizability coefficient (see Equation 7.1). These analyses used the ratings that the interviewers made before they compared their judgments and discussed their discrepancies. In the design of the analysis, (a) interviewers are nested within interviewer-pairs, (b) Soldiers are nested within interviewer pairs, and (c) Soldiers are crossed with interviewers within each interviewer pair. In other words, each Soldier was rated by only one interviewer pair (interviewers nested within interviewer pairs), but he or she was rated by both interviewers within the pair (Soldiers crossed with interviewers). Soldiers, interviewers, and interviewer pairs were treated as random effects. These agreement values were computed for the ratings *before* consensus. Interrater reliability *within* interviewer pair was also computed (see Equation 7.2). The residual variance in each equation is Soldiers-by-interviewers nested within interviewer-pair. The design of the interview precluded the computation of interrater reliability *across* interviewer pairs. Table 7.6 shows that the interviewer pairs tended to provide consistent (i.e., reliable) ratings for each scale and composite score.

Interviewer agreement across interviewer pairs

$$\rho^2 = \frac{\sigma^2_{sold:intPair} + \sigma^2_{intPair} - \frac{\sigma^2_{residual}}{n_{int:intPair}}}{\sigma^2_{sold:intPair} + \sigma^2_{intPair} - \frac{\sigma^2_{residual}}{n_{int:intPair}} + \frac{\sigma^2_{residual}}{n_{int:intPair}} + \frac{\sigma^2_{int:intPair}}{n_{int:intPair}}} \qquad (7.1)$$

Interviewer reliability within interviewer pairs

$$\rho^2 = \frac{\sigma^2_{sold:intPair}}{\sigma^2_{sold:intPair} + \frac{\sigma^2_{residual}}{n_{int:intPair}}} \qquad (7.2)$$

Table 7.6. Interview Interrater Pre-Consensus Agreement and Reliability Estimates

	Agreement across Interviewer Pairs		Reliability within Interviewer Pairs	
Scale	2 Raters	1 Rater	2 Raters	1 Rater
1. Adaptability	.73	.47	.75	.60
2. Military Presence	.72	.45	.75	.60
3. Level of Effort & Initiative on the Job	.70	.41	.72	.56
4. Level of Integrity & Discipline on the Job	.77	.55	.79	.65
5. Relating to and Supporting Peers	.70	.42	.72	.57
6. Leadership Skills/ Potential	.69	.39	.71	.55
7. Oral Communication Skill	.76	.52	.78	.64
8. Self-Management/Self-Directed Learning Skill	.77	.54	.79	.66
9. MOS/Occupation-Specific Knowledge and Skill	.81	.63	.82	.69
Composite Interview Score	.88	.77	.87	.78
Composite Interview Score Excluding MOS-Specific Ratings	.88	.76	87	.77

Note. n = 183 for *MOS/Occupation-Specific Knowledge and Skill.* Sample sizes for the other scales range from 938–942. The design of the interviews prevented the computation of interrater *reliability* estimates across interviewer pairs. The agreement values represent a lower-bound estimate of the interrater reliability estimates.

Validity Estimates

Table 7.7 shows the criterion-related validity for the interview using the total composite score that excluded the *MOS/Occupation-Specific Knowledge and Skill* rating. Four criteria were used in the validity computations: the observed performance composite, the expected future performance composite, the senior NCO potential rating, and the overall effectiveness rating. These validity estimates (corrected for range restriction in the predictor and for criterion unreliability) ranged from .25 to .27.

Table 7.7. Corrected and Raw Correlations between the Interview (Excludes MOS/Occupation-Specific Knowledge) and Criteria for E5 Soldiers

Criterion	*r*
Observed Performance Composite	.25 (.17*)
Expected Future Performance Composite	.26 (.15*)
Senior NCO Potential Rating	.27 (.17*)
Overall Effectiveness Rating	.26 (.16*)

Note. $n = 471–474$. Correlations corrected for indirect range restriction on the predictor and for criterion unreliability appear outside of the parentheses. Raw correlations appear inside parentheses.

* $p < .05$ (one-tailed).

Differential Prediction Analyses

Differential prediction analyses (see Chapter 4 for a description of these analyses) were performed to determine whether the interview-criterion prediction equation differed across gender or race. Only E5 Soldiers were used for the differential prediction analyses because only they had both interview scores and criterion ratings. The results of these analyses are shown in Table 7.8.[28]

Table 7.8 shows that the only significant effect is a gender effect of –0.29 for expected future performance (which is scored on a 1–7 scale). This means that, at the mean interview score, the predicted future performance score is 0.29 point lower for females than for males. This represents 0.30 of a standard deviation unit on the expected future performance scale. A common regression line would overpredict expected future performance of females compared to males; thus, the bias favors females. The other demographic main effects were very small and not statistically significant.

There were no significant race or gender differences in terms of the interview main effect. In other words, the validity estimates did not differ significantly by gender or race. The bottom of Table 7.8 shows the criterion-related validity coefficients.

[28] To ease interpretation of the unstandardized regression weights, interview scores were standardized within pay grade prior to conducting these analyses. The demographic variables were coded as follows for purposes of analysis: race (white = 0, black = 1); gender (male = 0, female = 1.)

Table 7.8. Differential Prediction Analyses for the Interview (Excludes MOS/Occupation-Specific Knowledge)

Criterion	Demographic Main Effect		Interview Score Main Effect			
			Gender		Race	
	Gender	Race	Male	Female	White	Black
MMR statistics						
Expected Future Performance Comp.	–0.29 *	0.03	0.14	0.17	0.20	0.04
Observed Performance Comp.	–0.06	0.00	0.13	0.26	0.16	0.09
Correlations with Interview Score						
Expected Future Performance Comp.			.14	.17	.19	.04
Observed Performance Comp.			.15	.28	.17	.11

Note. n_{Gender} = 469–472; n_{Race} = 401–403. n_{Males} = 407, $n_{Females}$ = 62, n_{White} = 280, n_{Black} = 119. The validity estimates (i.e., correlations with interview score) are uncorrected.

*$p < .05$ (two-tailed) for demographic main effect.

In summary, there is only one significant fairness issue with the semi-structured interview. The interview actually favors females because it overpredicts their expected future performance. There were no race effects in terms of fairness or mean group differences. Although there were no significant race effects, the near-zero validity coefficient for blacks (when predicting expected future performance) is a concern.

Fairness analyses could not be performed for E4 Soldiers because they did not receive performance ratings. There were no significant mean differences between subgroups, however.

Interviewer Evaluations

At each test site, after all interviews were conducted, the senior NCO interviewers completed a questionnaire that asked their opinions about the semi-structured interview. Participants used a 5-point scale ("not at all" to "a very great extent") to indicate their satisfaction with the various components of the interview. The data suggested the interviewers were generally satisfied with the interview and considered it to be at least moderately useful to the E5/E6 promotion process (see Table 7.9). The data suggested no major problems with the interview or the training. The interviewers were also encouraged to provide written feedback about the interview. Written comments were few, but they primarily addressed specific questions in the question bank.

The interviewers were also asked, "Should this structured interview supplement or replace the Promotion Board appearance?" Among the 40 interviewers who responded, 5 (13%) said the interview should replace the Board, 20 (50%) said it should supplement the Board, and 15 (37%) said the interview should not be used for promotion. Several interviewers thought the interview would be useful for NCO development. The most commonly mentioned benefit of the interview was that it assesses leadership ability. In the opinion of some of the interviewers, leadership ability receives insufficient attention in the current promotion system, so they were pleased to see a tool to assess it. On the other hand, some interviewers were concerned that the expected answers would eventually become known; thus, Soldiers could do well in the interview

by finding out the expected answers. Further, many interviewers thought that some useful aspects of the Promotion Board are absent from the interview: observing a Soldier under stress, giving points for awards and education, and assessing "general Soldiering."

Table 7.9. Evaluation Results for the Semi-Structured Interview

Components of the Interview	Percent Responding		
	Not at All/Slight Extent	Moderate Extent	Great Extent/Very Great Extent
1. This structured interview would provide useful information to the E5/E6 promotion process.	22.4	37.9	39.6
2. The training was sufficient preparation for conducting these interviews.	0.0	27.6	72.4
3. The definitions of the *Performance Areas* are clear and concise.	3.4	34.5	62.1
4. The Soldiers/interviewees understood the questions that were selected from the *Question Bank.*	8.8	21.1	70.1
5. The Soldiers/interviewees understood the questions that my interview pair developed.	1.6	17.2	81.1
6. Writing new questions was manageable.	0.0	19.0	81.0
7. The rating scale anchors were useful for evaluating interviewee responses to questions.	12.0	37.9	50.0
8. The Overall Average Score on the *Interview Summary Worksheets* accurately reflected my overall evaluation of the candidates' structured interview performances.	8.6	27.6	63.8

Note. $n = 57–58$.

Summary

The semi-structured interview obtained favorable results in the validation. It had moderate criterion-related validity, and it was well-received by the interviewers. Blacks performed as well as whites, and E4 females performed as well as E4 males. The only psychometric problems that it might have are that (a) although the difference was not significant, E5 females did not do as well as males; and (b) the interview scores' correlation with future performance for blacks was low (although not significantly lower than for whites). Further analysis would be required to draw any conclusions about these two psychometric issues.

The interview's main obstacles to implementation as part of a promotion system are practical. In the validation, the interviews lasted 30 minutes, and the interviewers completed 3–4 hours of training. The procedures, duration, and training could be modified to some extent to make the interview more acceptable for promotion. Alternatively, the interview could be useful in a training-and-development context, which is a view shared by several senior NCOs who served as interviewers.

CHAPTER 8: TEMPERAMENT INVENTORIES

Dan J. Putka
HumRRO

Robert N. Kilcullen and Leonard A.White
U.S. Army Research Institute for the Behavioral and Social Sciences

Overview

This chapter describes the validation of two self-report temperament-related inventories designed to capture information about Soldiers' personality traits and background considered relevant to the performance of 21st century NCOs. Elements of both measures are currently in operational use in various segments of the Army. The Assessment of Individual Motivation (AIM) serves as a supplemental screen for enlisted applicants who do not have a high school diploma and is being investigated for other Army uses. The Biographical Information Questionnaire (BIQ) comprises items from several biodata instruments that serve various purposes (e.g., screening Soldiers interested in joining the Special Forces). The following sections contain background information on each of these instruments, as well as information regarding their validation.

Assessment of Individual Motivation (AIM)

The AIM is a multidimensional forced choice inventory that reliably measures six temperament constructs: Dependability, Adjustment, Work Orientation, Leadership, Agreeableness, and Physical Conditioning (White & Young, 1998; Young, Heggestad, Rumsey, & White, 2000). Definitions for these constructs, as assessed by their respective AIM scales, are shown in Table 8.1. These constructs are closely related to several NCO21 KSAs, namely (a) *Need for Achievement*, (b) *Conscientiousness/Dependability*, (c) *Emotional Stability*, and (d) *Adaptability* (Knapp et al., 2002).

In the Army's Project A research, these constructs were measured by a self-report instrument called the Assessment of Background and Life Experiences (ABLE). Originally, there was much interest in using ABLE for enlisted personnel selection and classification decisions, but its proposed implementation was withdrawn largely due to concerns about its susceptibility to response distortion (i.e., faking; White & Young, 2001). Given these concerns, ARI developed the AIM to measure the performance-relevant constructs from ABLE with greater resistance to faking.

Development of the AIM

Over a 4-year period, seven developmental versions of AIM were administered to approximately 5,000 new Army recruits. Over several iterations, test forms were administered and refined until the prototype AIM form was finalized and evaluated in 1996.

The strategy for developing the AIM differed from that of the ABLE in several significant ways (White & Young, 1998; White, 2002). First, ABLE uses a forced-choice format to reduce item transparency and place constraints on faking. AIM items consist of four

statements (a tetrad) that may describe an examinee's past behavior in familiar situations. Two of these statements are worded positively (often indicating a high standing on the construct) and two are worded negatively (often indicating a low standing on the construct). For each item, respondents are asked to select the one statement (stem) which is *most like* them, and the one statement which is *least like* them. The version of the AIM used in this validation effort comprises 38 items. A quasi-ipsative scoring method is used to generate four construct scores for each item (i.e., one score for each stem). Scale scores are obtained by summing—across items—the scores for stems measuring the same construct.

Table 8.1. Definitions of Constructs Assessed by AIM Scales

Title	Definition
Work Orientation	The tendency to strive for excellence in the completion of work-related tasks. Persons high on this construct seek challenging work activities and set high standards for themselves. They consistently work hard to meet these high standards.
Adjustment	The tendency to have a uniformly positive affect. Persons high on this construct maintain a positive outlook on life, are free of excessive fears and worries, and have a feeling of self-control. They maintain their positive affect and self-control even when faced with stressful circumstances.
Agreeableness	The tendency to interact with others in a pleasant manner. Persons high on this construct get along and work well with others. They show kindness, while avoiding arguments and negative emotional outbursts directed at others.
Dependability	The tendency to respect and obey rules, regulations, and authority figures. Persons high on this construct are more likely to stay out of trouble in the workplace and avoid getting into difficulties with law enforcement officials.
Leadership	The tendency to seek out and enjoy being in leadership positions. Persons high on this scale are confident of their abilities and gravitate towards leadership roles in groups. They feel comfortable directing the activities of other people and are looked to for direction when group decisions have to be made.
Physical Conditioning	The tendency to seek out and participate in physically demanding activities. Persons high on this construct routinely participate in vigorous sports or exercise, and enjoy hard physical work.

Another important strategy in AIM's development was to create items that focused as much as possible on behaviors—thereby making them more like biodata. This contrasts with ABLE, which contains items relating to personal attitudes, affect, and traits. However, research from ABLE was very useful in identifying past experiences and behaviors linked to the target constructs, and therefore helped to guide ARI's development and revision of the AIM items. Further details on the development AIM are reported elsewhere (White & Young, 1998; Young et al., 2000).

Results

Data Preparation

Soldiers' responses to AIM items were carefully screened prior to conducting any validation analyses. AIM data were first reviewed for missing responses. We retained any individual who responded to at least 90% of AIM responses (69 out of 76). Of the 1,881 Soldiers who completed the

AIM, only 37 provided fewer than 69 responses. These 37 Soldiers were eliminated from all further analyses. The AIM data were also reviewed for evidence of patterned responding (e.g., a Soldier always choosing the first behavioral statement as "most like me"). Based on a careful review of the data by two psychologists, 11 Soldiers' AIM data were removed from further analyses. In sum, 1,835 Soldiers had usable AIM data for subsequent analysis.[29]

Descriptive Statistics

Table 8.2 presents coefficients alpha and intercorrelations for the AIM scores by pay grade. Although an estimate of internal consistency may be inappropriate given the AIM's partially ipsative scaling (Hicks, 1970), it may still be a useful heuristic for making comparisons among the AIM scores themselves. With the exception of coefficients alpha for the AIM Dependability scale among E5 and E6 Soldiers, all AIM scales had coefficients alpha greater than .60.

Table 8.2. AIM Score Intercorrelations and Reliability Estimates

Predictor	AIM Depend	AIM Adjust	AIM Work Or	AIM Agree	AIM Phy Cond	AIM Leader
E4 Soldiers						
AIM Dependability	(.67)					
AIM Adjustment	.37*	(.68)				
AIM Work Orientation	.42*	.34*	(.73)			
AIM Agreeableness	.55*	.47*	.39*	(.65)		
AIM Physical Conditioning	.32*	.32*	.43*	.29*	(.63)	
AIM Leadership	.25*	.40*	.60*	.21*	.10*	(.75)
E5 Soldiers						
AIM Dependability	(.57)					
AIM Adjustment	.30*	(.69)				
AIM Work Orientation	.33*	.31*	(.73)			
AIM Agreeableness	.52*	.43*	.30*	(.64)		
AIM Physical Conditioning	.22*	.24*	.35*	.27*	(.65)	
AIM Leadership	.19*	.34*	.58*	.12*	.03	(.74)
E6 Soldiers						
AIM Dependability	(.55)					
AIM Adjustment	.31*	(.70)				
AIM Work Orientation	.26*	.24*	(.69)			
AIM Agreeableness	.46*	.48*	.20*	(.64)		
AIM Physical Conditioning	.20*	.28*	.32*	.18*	(.61)	
AIM Leadership	.11*	.28*	.59*	.06	.01	(.70)

Note. n_{E4} = 434-435 ; n_{E5} = 860-861; n_{E6} = 537. Correlations are uncorrected. Internal consistency reliability estimates (coefficients alpha) are in parentheses.

*$p < .05$ (one-tailed).

[29] Some soldiers who had more than 10% of their AIM responses missing also exhibited patterned responding. Thus, the reported number of soldiers eliminated for missing data and the reported number of soldiers eliminated for patterned responding are not mutually exclusive.

Descriptive statistics for the six AIM scores, by subgroup (pay grade, race, gender, and CMF cluster) are presented in Tables 8.3 through 8.14. Raw and conditional effect sizes were calculated using methods described in Chapter 3. Tables 8.3 through 8.14 provide little evidence of large subgroup differences on any of the AIM scores.

Table 8.3. Subgroup Differences by Pay Grade, Gender, and Race for AIM Dependability

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	1.32	0.25	0.76	<.001	64	1.25	0.25	0.52	.006
Male	354	1.12	0.25			311	1.12	0.25		
Race										
Black	90	1.19	0.26	0.17	.153	87	1.17	0.24	-0.07	.608
White	290	1.14	0.26			288	1.19	0.26		
E5										
Gender										
Female	107	1.30	0.23	0.33	.002	88	1.26	0.24	0.17	.366
Male	751	1.23	0.22			659	1.22	0.22		
Race										
Black	238	1.23	0.22	0.03	.695	237	1.22	0.21	-0.16	.187
White	511	1.23	0.23			510	1.26	0.23		
E6										
Gender										
Female	57	1.37	0.18	0.55	<.001	46	1.33	0.16	0.30	.163
Male	479	1.25	0.21			415	1.26	0.22		
Race										
Black	175	1.30	0.20	0.27	.004	174	1.30	0.20	0.07	.653
White	290	1.24	0.22			287	1.29	0.22		
Grade										
E5	861	1.23	0.23	0.28	<.001	747	1.24	0.22	0.22	.037
E4	434	1.16	0.26			375	1.18	0.25		
E6	537	1.26	0.21	0.12	.041	461	1.30	0.21	0.26	.022
E5	861	1.23	0.23			747	1.24	0.22		
E6	537	1.26	0.21	0.39	<.001	461	1.30	0.21	0.45	<.001
E4	434	1.16	0.26			375	1.18	0.25		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Validity Estimates

Criterion-related validity was examined by computing zero-order correlations between the AIM scores and the criterion scores described in Chapter 3. Correlations were computed separately for E5 and E6 Soldiers, and differences between corresponding correlations (across pay grades) were tested for statistical significance. All correlations were corrected for direct range restriction and criterion unreliability per methods described earlier (cf. Chapter 4). Corrected and raw correlations are presented in Table 8.15.

The Work Orientation and Leadership scales exhibited high levels of validity for predicting observed and expected future performance among E5 Soldiers (Work Orientation: r = .40 and r = .46, respectively; Leadership: r = .33; r = .43, respectively), but significantly lower validity estimates for E6 Soldiers (Work Orientation: r = .13 and r = .17, respectively; Leadership: r = .09 and r = .12, respectively), though estimates for Work Orientation were still significantly greater than zero.

Such differences in the E5 and E6 validity estimates may indicate that temperament-related variation in performance is less critical to successful performance at the E6 level, relative to the E5 level, or simply that most E6 Soldiers possess the requisite levels of the traits assessed by the AIM. Although the means and standard deviations of E5 and E6 Soldiers on the AIM scales tend to be quite similar, it is quite possible that if temperament is less critical to E6 performance, then—all other things being equal—lower levels of temperament among E6 Soldiers would be sufficient for performing their jobs successfully. Thus, although the E5 and E6 means on the AIM scales are similar, more E6s may fall within the range of temperament that is sufficient for performance at their given pay grade, effectively attenuating the E6 validity estimates relative to the E5 estimates for the AIM scales.

Validity estimates for the Dependability and Physical Conditioning scales were low but statistically significant among E5 Soldiers for both observed (.17 for Dependability, .15 for Physical Conditioning) and expected future performance (.21 for Dependability, .16 for Physical Conditioning). Among E6 Soldiers, however, these scales showed little validity (r = -.02 and r = .03 for observed performance, respectively; r = .02 and r = .06 for expected performance, respectively). Although these differences between E5 and E6 correlations appear sizable, they are not statistically significant.

Table 8.4. Differences between CMF Clusters for AIM Dependability

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	67	54	1.27	1.25	0.23	0.24	—	0.07	-0.47	-0.18	.	-0.59 *
2. INT	21	20	1.25	1.27	0.28	0.27	-0.05	—	-0.54	-0.25	.	-0.66
3. CBO	169	154	1.10	1.14	0.25	0.25	-0.63**	-0.58*	—	0.29	.	-0.12
4. LOG	138	116	1.17	1.21	0.28	0.25	-0.36**	-0.30	0.27*	—	.	-0.41
5. CPA	12	10	1.17	1.13	0.24	0.19	.	.	.	.	—	.
6. COM	24	21	1.11	1.10	0.25	0.27	-0.58**	-0.53	0.05	-0.22	.	—
Overall	434		1.16		0.26							
E5 Soldiers												
1. ADM	81	68	1.29	1.32	0.24	0.22	—	-0.31	-0.55	-0.30 *	-0.33	-0.62 **
2. INT	37	35	1.23	1.25	0.27	0.27	-0.25	—	-0.25	0.01	-0.02	-0.31
3. CBO	324	280	1.21	1.19	0.22	0.22	-0.35**	-0.10	—	0.25	0.23	-0.07
4. LOG	280	240	1.24	1.25	0.22	0.22	-0.22	0.03	0.13	—	-0.03	-0.32
5. CPA	79	73	1.25	1.24	0.22	0.23	-0.16	0.09	0.19	0.06	—	-0.29
6. COM	58	51	1.22	1.18	0.23	0.22	-0.30	-0.05	0.05	-0.08	-0.14	—
Overall	861		1.23		0.23							
E6 Soldiers												
1. ADM	59	50	1.31	1.32	0.17	0.18	—	.	-0.43	-0.04	-0.29	0.02
2. INT	20	15	1.30	1.34	0.24	0.22	-0.04	—	.	.	.	.
3. CBO	199	174	1.23	1.23	0.21	0.21	-0.39**	-0.36	—	0.40	0.15	0.45
4. LOG	165	142	1.28	1.31	0.23	0.23	-0.16	-0.12	0.24*	—	-0.25	0.05
5. CPA	65	55	1.24	1.26	0.21	0.22	-0.33*	-0.30	0.06	-0.18	—	0.30
6. COM	29	25	1.34	1.32	0.18	0.17	0.14	0.18	0.53**	0.29	0.47 *	—
Overall	537		1.26		0.21							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.5. Subgroup Differences by Pay Grade, Gender, and Race for AIM Adjustment

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	1.17	0.23	-0.04	.749	64	1.12	0.24	-0.28	.192
Male	355	1.18	0.21			311	1.18	0.21		
Race										
Black	90	1.17	0.22	-0.03	.826	87	1.14	0.22	-0.09	.585
White	290	1.17	0.22			288	1.16	0.22		
E5										
Gender										
Female	107	1.13	0.24	-0.36	<.001	88	1.06	0.23	-0.64	<.001
Male	751	1.21	0.22			659	1.20	0.22		
Race										
Black	238	1.21	0.21	0.10	.184	237	1.14	0.21	0.12	.316
White	511	1.19	0.23			510	1.11	0.23		
E6										
Gender										
Female	57	1.18	0.20	-0.20	.153	46	1.10	0.20	-0.64	.003
Male	479	1.22	0.21			415	1.23	0.20		
Race										
Black	175	1.25	0.20	0.22	.018	174	1.20	0.19	0.31	.031
White	290	1.20	0.21			287	1.13	0.21		
Grade										
E5	861	1.20	0.22	0.12	.013	747	1.13	0.22	-0.10	.408
E4	435	1.17	0.22			375	1.15	0.22		
E6	537	1.22	0.21	0.07	.234	461	1.16	0.20	0.17	.124
E5	861	1.20	0.22			747	1.13	0.22		
E6	537	1.22	0.21	0.20	.001	461	1.16	0.20	0.07	.573
E4	435	1.17	0.22			375	1.15	0.22		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed t-tests of differences between subgroup means.

Table 8.6. Differences between CMF Clusters for AIM Adjustment

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	67	54	1.22	1.20	0.24	0.25	—	-0.32	-0.87 *	-0.14	.	-0.14
2. INT	21	20	1.20	1.13	0.25	0.26	-0.10	—	-0.55	0.18	.	0.18
3. CBO	169	154	1.15	1.01	0.20	0.20	-0.33 *	-0.23	—	0.72	.	0.73
4. LOG	139	116	1.17	1.17	0.22	0.22	-0.22	-0.12	0.11	—	.	0.00
5. CPA	12	10	1.22	1.21	0.22	0.22	.	.	.	.	—	.
6. COM	24	21	1.21	1.17	0.14	0.16	-0.03	0.07	0.30	0.19	.	—
Overall	435		1.17		0.22							
E5 Soldiers												
1. ADM	81	68	1.20	1.19	0.20	0.19	—	-0.55	-0.71	-0.09	-0.01	-0.36
2. INT	37	35	1.17	1.07	0.26	0.22	-0.15	—	-0.16	0.46	0.55	0.19
3. CBO	324	280	1.21	1.03	0.23	0.23	0.02	0.17	—	0.62	0.70	0.35
4. LOG	280	240	1.20	1.17	0.22	0.23	-0.00	0.15	-0.02	—	0.09	-0.27
5. CPA	79	73	1.18	1.19	0.20	0.21	-0.12	0.03	-0.14	-0.11	—	-0.35
6. COM	58	51	1.16	1.11	0.22	0.19	-0.19	-0.04	-0.21	-0.19	-0.07	—
Overall	861		1.20		0.22							
E6 Soldiers												
1. ADM	59	50	1.25	1.22	0.20	0.19	—	.	-0.71	-0.20	0.22	-0.26
2. INT	20	15	1.17	1.10	0.24	0.25	-0.36	—	.	.	.	.
3. CBO	199	174	1.21	1.07	0.22	0.20	-0.17	0.19	—	0.52	0.94 *	0.45
4. LOG	165	142	1.21	1.18	0.19	0.19	-0.19	0.17	-0.03	—	0.42	-0.06
5. CPA	65	55	1.23	1.26	0.22	0.21	-0.08	0.27	0.08	0.11	—	-0.48
6. COM	29	25	1.25	1.16	0.27	0.27	0.03	0.39	0.20	0.23	0.12	—
Overall	537		1.22		0.21							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw effect size statistic; Con = Conditional effect size statistic. Raw effect sizes calculated as (mean of higher-numbered category – mean of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.7. Subgroup Differences by Pay Grade, Gender, and Race for AIM Work Orientation

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	1.28	0.24	0.30	.020	64	1.25	0.24	0.31	.111
Male	355	1.20	0.26			311	1.17	0.26		
Race										
Black	90	1.19	0.22	-0.06	.615	87	1.19	0.21	-0.11	.460
White	290	1.21	0.27			288	1.22	0.27		
E5										
Gender										
Female	107	1.30	0.21	0.06	.551	88	1.29	0.21	0.12	.490
Male	751	1.28	0.25			659	1.26	0.25		
Race										
Black	238	1.25	0.23	-0.17	.032	237	1.25	0.23	-0.22	.068
White	511	1.29	0.25			510	1.31	0.25		
E6										
Gender										
Female	57	1.34	0.20	0.02	.871	46	1.35	0.22	0.04	.871
Male	479	1.33	0.22			415	1.34	0.22		
Race										
Black	175	1.32	0.20	-0.09	.358	174	1.34	0.21	-0.08	.590
White	290	1.34	0.23			287	1.35	0.23		
Grade										
E5	861	1.29	0.24	0.27	<.001	747	1.28	0.24	0.28	.010
E4	435	1.22	0.26			375	1.21	0.26		
E6	537	1.33	0.22	0.19	.002	461	1.35	0.22	0.27	.012
E5	861	1.29	0.24			747	1.28	0.24		
E6	537	1.33	0.22	0.45	<.001	461	1.35	0.22	0.54	<.001
E4	435	1.22	0.26			375	1.21	0.26		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.8. Differences between CMF Clusters for AIM Work Orientation

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	67	54	1.27	1.24	0.20	0.20	—	0.01	-0.15	-0.00	.	-0.65 *
2. INT	21	20	1.23	1.25	0.25	0.22	-0.15	—	-0.16	-0.01	.	-0.65
3. CBO	169	154	1.19	1.21	0.26	0.26	-0.30**	-0.15	—	0.15	.	-0.49
4. LOG	139	116	1.24	1.24	0.27	0.27	-0.11	0.04	0.19	—	.	-0.64 *
5. CPA	12	10	1.22	1.22	0.34	0.32	.	.	.	.	—	.
6. COM	24	21	1.15	1.08	0.26	0.27	-0.46*	-0.30	-0.15	-0.35	.	—
Overall	435		1.22		0.26							
E5 Soldiers												
1. ADM	81	68	1.30	1.31	0.24	0.23	—	0.11	-0.19	-0.17	-0.16	-0.38
2. INT	37	35	1.31	1.34	0.21	0.20	0.05	—	-0.29	-0.27	-0.27	-0.48
3. CBO	324	280	1.29	1.27	0.24	0.24	-0.02	-0.07	—	0.02	0.02	-0.19
4. LOG	280	240	1.28	1.27	0.25	0.25	-0.08	-0.13	-0.06	—	0.00	-0.21
5. CPA	79	73	1.26	1.27	0.26	0.26	-0.15	-0.20	-0.13	-0.07	—	-0.21
6. COM	58	51	1.29	1.22	0.24	0.24	-0.03	-0.08	-0.01	0.05	0.12	—
Overall	861		1.29		0.24							
E6 Soldiers												
1. ADM	59	50	1.29	1.34	0.24	0.20	—	.	-0.09	-0.03	0.08	-0.29
2. INT	20	15	1.39	1.45	0.23	0.23	0.42	—	.	.	.	.
3. CBO	199	174	1.33	1.32	0.22	0.21	0.17	-0.24	—	0.06	0.17	-0.20
4. LOG	165	142	1.34	1.33	0.20	0.21	0.20	-0.22	0.03	—	0.11	-0.26
5. CPA	65	55	1.32	1.36	0.26	0.26	0.10	-0.31	-0.07	-0.10	—	-0.37
6. COM	29	25	1.34	1.28	0.23	0.25	0.20	-0.22	0.02	-0.01	0.09	—
Overall	537		1.33		0.22							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.9. Subgroup Differences by Pay Grade, Gender, and Race for AIM Agreeableness

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	1.25	0.22	0.18	.152	64	1.24	0.22	0.06	.740
Male	355	1.21	0.24			311	1.22	0.24		
Race										
Black	90	1.22	0.23	0.09	.457	87	1.23	0.24	-0.02	.880
White	290	1.20	0.25			288	1.23	0.24		
E5										
Gender										
Female	107	1.25	0.22	-0.03	.755	88	1.26	0.22	-0.05	.770
Male	750	1.26	0.23			658	1.27	0.24		
Race										
Black	238	1.26	0.22	0.05	.509	237	1.26	0.22	-0.05	.670
White	510	1.25	0.24			509	1.27	0.24		
E6										
Gender										
Female	57	1.28	0.23	-0.02	.865	46	1.29	0.25	-0.13	.560
Male	479	1.29	0.22			415	1.31	0.22		
Race										
Black	175	1.33	0.21	0.34	<.001	174	1.33	0.21	0.28	.050
White	290	1.26	0.23			287	1.27	0.23		
Grade										
E5	860	1.26	0.23	0.18	<.001	746	1.26	0.23	0.14	.220
E4	435	1.21	0.24			375	1.23	0.24		
E6	537	1.29	0.22	0.12	.041	461	1.30	0.22	0.17	.130
E5	860	1.26	0.23			746	1.26	0.23		
E6	537	1.29	0.22	0.30	<.001	461	1.30	0.22	0.30	.020
E4	435	1.21	0.24			375	1.23	0.24		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.10. Differences between CMF Clusters for AIM Agreeableness

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	67	54	1.30	1.29	0.22	0.23	—	-0.16	-0.15	-0.31	.	-0.44
2. INT	21	20	1.26	1.25	0.27	0.28	-0.18	—	0.01	-0.16	.	-0.29
3. CBO	169	154	1.17	1.25	0.24	0.24	-0.54**	-0.36	—	-0.17	.	-0.30
4. LOG	139	116	1.22	1.21	0.23	0.23	-0.34*	-0.16	0.20	—	.	-0.13
5. CPA	12	10	1.20	1.18	0.25	0.25	.	.	.	.	—	.
6. COM	24	21	1.21	1.18	0.25	0.28	-0.38	-0.19	0.16	-0.03	.	—
Overall	435		1.21		0.24							
E5 Soldiers												
1. ADM	81	68	1.29	1.28	0.22	0.22	—	-0.16	0.07	-0.09	-0.14	-0.27
2. INT	37	35	1.29	1.25	0.24	0.24	0.01	—	0.23	0.07	0.02	-0.11
3. CBO	323	279	1.24	1.30	0.23	0.23	-0.23*	-0.25	—	-0.16	-0.22	-0.34
4. LOG	280	240	1.27	1.26	0.24	0.24	-0.09	-0.10	0.15	—	-0.06	-0.18
5. CPA	79	73	1.24	1.25	0.23	0.24	-0.22	-0.23	0.02	-0.13	—	-0.12
6. COM	58	51	1.26	1.22	0.23	0.23	-0.12	-0.13	0.12	-0.03	0.10	—
Overall	860		1.26		0.23							
E6 Soldiers												
1. ADM	59	50	1.32	1.31	0.22	0.23	—	.	0.18	-0.20	-0.05	0.17
2. INT	20	15	1.26	1.24	0.19	0.22	-0.30	—	.	.	.	.
3. CBO	199	174	1.27	1.35	0.24	0.23	-0.25	0.05	—	-0.38	-0.23	-0.01
4. LOG	165	142	1.28	1.27	0.22	0.22	-0.20	0.10	0.05	—	0.15	0.37
5. CPA	65	55	1.28	1.30	0.21	0.20	-0.19	0.11	0.05	0.01	—	0.22
6. COM	29	25	1.39	1.35	0.20	0.20	0.30	0.60*	0.55**	0.50*	0.50*	—
Overall	537		1.29		0.22							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.11. Subgroup Differences by Pay Grade, Gender, and Race for AIM Physical Conditioning

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	1.21	0.34	0.14	.287	64	1.05	0.32	-0.26	.206
Male	355	1.17	0.28			311	1.12	0.28		
Race										
Black	90	1.17	0.27	0.00	.970	87	1.05	0.26	-0.24	.123
White	290	1.17	0.30			288	1.12	0.29		
E5										
Gender										
Female	107	1.23	0.28	0.01	.948	88	1.16	0.29	-0.25	.173
Male	750	1.23	0.29			658	1.23	0.29		
Race										
Black	238	1.23	0.27	0.03	.737	237	1.17	0.27	-0.17	.143
White	510	1.22	0.30			509	1.22	0.30		
E6										
Gender										
Female	57	1.25	0.23	0.09	.529	46	1.19	0.23	-0.19	.370
Male	479	1.23	0.27			415	1.24	0.27		
Race										
Black	175	1.24	0.24	0.07	.418	174	1.19	0.24	-0.18	.215
White	290	1.22	0.28			287	1.24	0.28		
Grade										
E5	860	1.23	0.29	0.18	<.001	746	1.19	0.29	0.38	<.001
E4	435	1.18	0.29			375	1.08	0.29		
E6	537	1.23	0.27	-0.01	.895	461	1.22	0.26	0.09	.414
E5	860	1.23	0.29			746	1.19	0.29		
E6	537	1.23	0.27	0.18	.003	461	1.22	0.26	0.47	<.001
E4	435	1.18	0.29			375	1.08	0.29		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.12. Differences between CMF Clusters for AIM Physical Conditioning

CMFC	n		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	67	54	1.26	1.25	0.27	0.29	—	-0.56	-0.89 *	-0.25	.	-0.87 **
2. INT	21	20	1.15	1.09	0.28	0.27	-0.36	—	-0.34	0.31	.	-0.31
3. CBO	169	154	1.18	0.99	0.28	0.28	-0.27*	0.09	—	0.64	.	0.02
4. LOG	139	116	1.18	1.17	0.30	0.29	-0.27	0.10	0.00	—	.	-0.62 *
5. CPA	12	10	1.03	1.01	0.38	0.38	.	.	.	.	—	.
6. COM	24	21	1.08	1.00	0.28	0.27	-0.61**	-0.24	-0.34	-0.34 **	.	—
Overall	435		1.18		0.29							
E5 Soldiers												
1. ADM	81	68	1.23	1.23	0.30	0.30	—	-0.02	-0.61	-0.07	0.14	-0.15
2. INT	37	35	1.30	1.22	0.27	0.28	0.27	—	-0.59	-0.05	0.16	-0.13
3. CBO	323	279	1.24	1.05	0.27	0.28	0.05	-0.22	—	0.54	0.76	0.46
4. LOG	280	240	1.21	1.21	0.29	0.29	-0.06	-0.33*	-0.12	—	0.21	-0.08
5. CPA	79	73	1.22	1.27	0.33	0.34	-0.01	-0.28	-0.06	0.05	—	-0.30
6. COM	58	51	1.25	1.18	0.27	0.27	0.08	-0.19	0.03	0.14	0.09	—
Overall	860		1.23		0.29							
E6 Soldiers												
1. ADM	59	50	1.24	1.28	0.32	0.28	—	.	-0.86	-0.25	-0.02	-0.15
2. INT	20	15	1.30	1.25	0.28	0.28	0.25	—	.	.	.	.
3. CBO	199	174	1.22	1.05	0.27	0.26	-0.05	-0.30	—	0.60	0.84	0.71 *
4. LOG	165	142	1.22	1.21	0.26	0.26	-0.07	-0.32	-0.02	—	0.24	0.11
5. CPA	65	55	1.23	1.28	0.27	0.28	-0.02	-0.27	0.03	0.05	—	-0.13
6. COM	29	25	1.27	1.24	0.25	0.26	0.11	-0.14	0.17	0.19	0.13	—
Overall	537		1.23		0.27							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.13. Subgroup Differences by Pay Grade, Gender, and Race for AIM Leadership

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	1.25	0.27	0.18	.169	64	1.33	0.26	0.51	.008
Male	355	1.20	0.24			311	1.20	0.24		
Race										
Black	90	1.20	0.21	-0.03	.779	87	1.28	0.20	0.12	.374
White	290	1.21	0.26			288	1.25	0.26		
E5										
Gender										
Female	107	1.25	0.22	-0.10	.332	88	1.29	0.22	0.11	.542
Male	751	1.28	0.23			659	1.26	0.23		
Race										
Black	238	1.25	0.19	-0.13	.073	237	1.28	0.19	0.03	.783
White	511	1.28	0.24			510	1.27	0.24		
E6										
Gender										
Female	57	1.26	0.21	-0.32	.025	46	1.32	0.23	-0.09	.676
Male	479	1.33	0.21			415	1.34	0.21		
Race										
Black	175	1.29	0.21	-0.26	.006	174	1.33	0.21	0.00	.998
White	290	1.35	0.21			287	1.33	0.21		
Grade										
E5	861	1.27	0.23	0.26	<.001	747	1.28	0.23	0.05	.619
E4	435	1.21	0.24			375	1.26	0.25		
E6	537	1.32	0.21	0.21	<.001	461	1.33	0.21	0.23	.041
E5	861	1.27	0.23			747	1.28	0.23		
E6	537	1.32	0.21	0.45	<.001	461	1.33	0.21	0.26	.030
E4	435	1.21	0.24			375	1.26	0.25		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.14. Differences between CMF Clusters for AIM Leadership

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	67	54	1.21	1.19	0.27	0.27	—	0.63 *	0.43	0.11	.	-0.05
2. INT	21	20	1.31	1.35	0.30	0.29	0.40	—	-0.19	-0.52	.	-0.67
3. CBO	169	154	1.21	1.30	0.24	0.24	-0.01	-0.41	—	-0.32	.	-0.48
4. LOG	139	116	1.20	1.22	0.22	0.23	-0.05	-0.45	-0.04	—	.	-0.15
5. CPA	12	10	1.32	1.34	0.31	0.26	.	.	.	.	—	.
6. COM	24	21	1.22	1.18	0.24	0.25	0.06	-0.34	0.07	0.11	.	—
Overall	435		1.21		0.24							
E5 Soldiers												
1. ADM	81	68	1.27	1.27	0.23	0.20	—	0.17	0.26	-0.12	0.15	-0.20
2. INT	37	35	1.27	1.31	0.21	0.22	0.01	—	0.09	-0.29	-0.01	-0.37
3. CBO	324	280	1.29	1.33	0.23	0.23	0.09	0.08	—	-0.38	-0.11	-0.46
4. LOG	280	240	1.26	1.24	0.23	0.23	-0.06	-0.07	-0.15	—	0.27	-0.08
5. CPA	79	73	1.28	1.30	0.20	0.20	0.03	0.02	-0.06	0.09	—	-0.36
6. COM	58	51	1.28	1.22	0.24	0.22	0.03	0.02	-0.06	0.09	-0.00	—
Overall	861		1.27		0.23							
E6 Soldiers												
1. ADM	59	50	1.27	1.27	0.21	0.20	—	.	0.52	0.06	0.24	0.15
2. INT	20	15	1.40	1.42	0.22	0.25	0.59*	—	.	.	.	.
3. CBO	199	174	1.35	1.38	0.22	0.22	0.35*	-0.24	—	-0.47	-0.28	-0.38
4. LOG	165	142	1.30	1.28	0.18	0.18	0.11	-0.48	-0.23*	—	0.18	0.09
5. CPA	65	55	1.31	1.32	0.22	0.24	0.16	-0.43	-0.18	0.05	—	-0.09
6. COM	29	25	1.36	1.30	0.23	0.21	0.43	-0.16	0.08	0.31	0.26	—
Overall	537		1.32		0.21							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.15. Corrected and Raw Correlations between AIM Scores and Criteria for E5 and E6 Soldiers

Criterion	Predictor					
	AIM Dependability	AIM Adjustment	AIM Work Orientation	AIM Agreeableness	AIM Physical Conditioning	AIM Leadership
E5 Soldiers						
Observed Performance Composite	.17 (.11*)	.08 (.06)	.40 (.28$_a$*)	.02 (.01)	.15 (.11*)	.33 (.22$_a$*)
Expected Future Performance Composite	.21 (.12*)	.08 (.05)	.46 (.28$_a$*)	-.02 (-.01)	.16 (.10*)	.43 (.26$_a$*)
Senior NCO Potential Rating	.12 (.07*)	.03 (.02)	.44 (.28$_a$*)	.03 (.02)	.17 (.11*)	.35 (.22*)
Overall Effectiveness Rating	.13 (.07$_a$*)	.05 (.03)	.40 (.25*)	.05 (.03)	.12 (.08*)	.32 (.20*)
E6 Soldiers						
Observed Performance Composite	-.02 (-.01)	.10 (.07)	.13 (.09*)	-.01 (-.01)	.03 (.02)	.09 (.06)
Expected Future Performance Composite	.02 (.01)	.19 (.12*)	.17 (.11*)	.03 (.02)	.06 (.04)	.12 (.08)
Senior NCO Potential Rating	-.08 (-.05)	.08 (.06)	.21 (.14*)	-.11 (-.08)	.08 (.05)	.16 (.10*)
Overall Effectiveness Rating	-.09 (-.06)	.10 (.07)	.20 (.13*)	-.07 (-.05)	.08 (.05)	.16 (.11*)

Note. n_{E5} = 593-598; n_{E6} = 384-388. Correlations corrected criterion unreliability and for direct range restriction on the predictor appear outside of the parentheses. Raw correlations appear inside parentheses. The "a" subscripts on E5 correlations indicate that corresponding E5 and E6 correlations were significantly different from each other, $p < .05$ (two-tailed).

* $p < .05$ (one-tailed).

Lastly, the Adjustment and Agreeableness scores exhibited little to no validity for predicting the observed performance of E5 and E6 Soldiers. Nevertheless, the validity of the Adjustment score among E6 Soldiers was significant when predicting expected future performance ($r = .19$).

Differential Prediction Analyses

Table 8.16 presents the results of differential prediction analyses for AIM scores by pay grade and criterion, examining gender and race as the demographic variables of interest.[30] Overall, the results provide little evidence of differential prediction (i.e., slope bias). In the cases where differential prediction was evident, the better prediction appeared to be for the minority group. For example, the Adjustment score was more predictive of expected future performance for black E6 Soldiers ($b = 0.29$) than for white E6 Soldiers ($b = 0.05$).

Although no evidence of intercept bias emerged for race-based comparisons, evidence of intercept bias did emerge for gender-based comparisons when predicting expected future performance. Specifically, women had expected performance scores that were roughly 0.34 to 0.47 point lower than men (at mean levels of the AIM scores). These findings suggest that the AIM scores would tend to *overpredict* females' expected future performance if a common regression equation were used.

AIM Summary

Several AIM scores showed promise as predictors for future E4-to-E5 NCO promotion decisions (in particular Work Orientation and Leadership), but they exhibited less promise as predictors for future E5-to-E6 NCO promotion decisions. Such validity differences may indicate that temperament-related attributes are less critical to successful NCO performance at the E6 level than at the E5 level, or simply that most E6 Soldiers possess the requisite levels of the traits assessed by the AIM.

Although analyses revealed few differences among subgroups on the AIM scores, there was evidence of intercept bias for gender (females' performance being overpredicted) when predicting expected future performance. Nevertheless, little evidence emerged that suggested AIM scores (in general) would be differentially predictive of future NCO performance.

Biographical Information Questionnaire (BIQ)

The BIQ measures eight temperament constructs important to effective NCO performance: Hostility to Authority, Manipulativeness, Social Maturity, Tolerance for Ambiguity, Openness, Emergent Leadership, Social Perceptiveness, and Interpersonal Skill. Descriptions of the BIQ scales reflecting each of these constructs are shown in Table 8.17. These constructs are closely related to

[30] All AIM scores were standardized within pay grade to ease interpretation of the unstandardized regression weights prior to conducting these analyses. The demographic variables were coded as follows for purposes of analysis: race (white = 0, black = 1), gender (male = 0, female = 1).

several NCO21 KSAs: (a) *Conscientiousness/Dependability*, (b) *Level of Integrity and Discipline on the Job*, and (c) *Adherence to Regulations, Policies, and Procedures* (Knapp et al., 2002).

Table 8.16. Differential Prediction Analyses for AIM Scores

Criterion/Predictor	Demographic Main Effect		AIM Score Main Effect				*r*			
			Gender		Race		Gender		Race	
	Gender	Race	M	F	W	B	M	F	W	B
			Observed Performance Composite							
E5 Soldiers										
AIM Dependability	-.16	.00	.11	.03	.04	.22$_a$	**.12**	.03	.04	**.26**
AIM Adjustment	-.12	-.01	.04	.05	.02	.07	.05	.05	.02	.08
AIM Work Orientation	-.19	.02	.21	.48$_a$	.21	.26	**.26**	**.45**	**.24**	**.30**
AIM Agreeableness	-.15	-.02	.01	.01	-.06	.15$_a$	.01	.01	-.07	**.18**
AIM Physical Conditioning	-.15	-.00	.11	-.03	.03	.18	**.13**	-.03	.04	**.21**
AIM Leadership	-.18	-.01	.17	.40	.20	.20	**.21**	**.37**	**.24**	**.21**
E6 Soldiers										
AIM Dependability	-.08	-.11	.00	-.08	.04	-.08	.00	-.07	.06	-.09
AIM Adjustment	-.04	-.16	.03	.28	.04	.14	.04	**.32**	.06	**.16**
AIM Work Orientation	-.13	-.12	.07	.12	.08	.06	**.09**	.13	**.12**	.06
AIM Agreeableness	-.12	-.14	-.01	.01	-.05	.11	-.01	.02	-.07	.11
AIM Physical Conditioning	-.12	-.12	.02	.05	.02	.01	.02	.05	.02	.01
AIM Leadership	-.13	-.11	.05	-.04	.07	-.02	.07	-.05	.09	-.02
			Expected Future Performance Composite							
E5 Soldiers										
AIM Dependability	-.38*	.02	.14	.05	.05	.20	**.14**	.05	.05	**.21**
AIM Adjustment	-.35*	.00	.03	.02	.02	.09	.03	.03	.03	.09
AIM Work Orientation	-.39*	.03	.26	.42	.29	.22	**.27**	**.38**	**.29**	**.22**
AIM Agreeableness	-.37*	.00	-.02	.03	-.08	.09	-.02	.03	-.08	.10
AIM Physical Conditioning	-.37*	.01	.12	-.04	.05	.15	**.12**	-.04	.05	**.15**
AIM Leadership	-.37*	.02	.25	.33	.32	.17	**.25**	**.29**	**.33**	**.16**
E6 Soldiers										
AIM Dependability	-.36	-.11	.04	-.13	.09$_a$	-.13	.05	-.08	**.12**	-.11
AIM Adjustment	-.34*	-.19	.08	.34	.05	.29$_a$	**.10**	.26	.06	**.25**
AIM Work Orientation	-.44*	-.12	.10	.10	.08	.11	**.11**	.07	.09	.09
AIM Agreeableness	-.44*	-.15	.03	-.13	-.04	.12	.03	-.10	-.05	.10
AIM Physical Conditioning	-.44*	-.13	.03	.13	.00	.05	.03	.09	.00	.04
AIM Leadership	-.47*	-.11	.09	-.21	.09	.03	**.11**	-.18	.10	.03

Note. Regression analysis sample sizes: $n_{E5\ Gender}$ = 591-596; $n_{E5\ Race}$ = 513-515; $n_{E6\ Gender}$ = 383-387; $n_{E6\ Race}$ = 336-340. Smaller sample sizes underlie the reported correlations because they were calculated for each subgroup separately. Correlations are uncorrected. Bolded correlations are statistically significant, $p < .05$ (one-tailed).

*$p < .05$ (two-tailed).

Table 8.17. BIQ Scale Descriptions

Title	Definition
Tolerance for Ambiguity	This scale measures a person's preference for work environments in which the problems (and potential solutions) are unstructured and ill-defined. Those with high tolerance for ambiguity are comfortable working in rapidly changing work environments. Individuals scoring low prefer highly structured and predictable work settings.
Openness	This scale measures the degree to which a person is open to new ideas and experiences. High scorers on this scale are curious, imaginative, have broad interests, and enjoy learning new things. Individuals low in openness dislike extensive thought and contemplation and tend to be set in their ways of doing things.
Hostility to Authority	The degree to which a person respects and is willing to follow legitimate authority figures. High scorers are expressively angered by authority figures and may actively disregard their instructions and policies. Low scorers accept directives from superiors and easily adapt to structured work environments.
Manipulativeness	The degree to which the individual is straightforward and open in his/her interpersonal relationships. Those scoring high in this scale routinely use deception, lies, and short cuts in dealing with others. They are prone to treating others as objects to be used for personal gain and gratification. Low scoring individuals tend to be sincere, aboveboard and straightforward when interacting with others.
Social Maturity	A willingness to follow societal rules and regulations. High scorers tend to be law-abiding and respectful of the rights and property of others. They willingly conform to societal laws, customs, and expectations. Low scorers are highly rebellious and have a history of violating rules and norms.
Social Perceptiveness	This scale measures the degree to which a person can discern and recognize others' emotions and likely behaviors in interpersonal situations. Persons high in social insight are good at understanding others' motives and are less likely to be "caught off guard" by unexpected interpersonal behaviors.
Interpersonal Skill	This scale measures the degree to which a person establishes smooth and effective interpersonal relationships with others. Interpersonally skilled individuals are good listeners, behave diplomatically, and get along well with others. Persons with low scores on this measure have difficulty working with others and may intentionally or unconsciously promote interpersonal conflict and cause hurt feelings.
Emergent Leadership	The scale measures the degree to which a person takes on leadership roles in groups and in his or her interactions with others. High scorers on this scale are looked to for direction and guidance when group decisions are made and readily take on leadership roles.

Instrument Description

Previous research has shown that biodata scales can be used to measure personality constructs, have higher criterion-related validity, and are less easily faked than traditional self-report personality assessments (e.g., Kilcullen, White, Mumford, & Mack, 1995). The 156 self-report items that constitute the BIQ reflect prior behaviors and reactions to specific life events indicative of the targeted psychological constructs. BIQ items were drawn from existing biodata instruments the Army has used for operational and research purposes.

Items measuring Hostility to Authority, Manipulativeness, and Social Maturity were drawn from the Army's Assessment of Right Conduct (ARC). These three scales have been related to delinquency criteria and are being used for operational screening and assessment in the Army. Previous research has linked these attributes to (a) completion of the Special Forces Assessment and Selection (SFAS) course and (b) a lower incidence of disciplinary infractions among NCO and first-term enlisted personnel (e.g., Kilcullen, Mael, Goodwin, & Zazanis, 1999).

Items measuring Tolerance for Ambiguity and Openness were drawn from a biodata instrument that has been used to measure adaptability. In previous research, these scales were related to the performance of Special Forces in Robin Sage, a military exercise consisting of ambiguous and unforeseen dilemmas designed to mimic the Special Forces operational environment (Kilcullen, Chen, Zazanis, Carpenter, & Goodwin, 1999). In this exercise, the team leader's Tolerance for Ambiguity and Openness scores were primary determinants of the SF team's ability to overcome these challenges and perform successfully.

Items for the remaining three biodata scales—Emergent Leadership, Social Perceptiveness, and Interpersonal Skill—were drawn from ARI-sponsored research involving determinants of military and civilian leadership effectiveness. In research with Army civilians, these measures, along with individual differences in supervisors' Tolerance for Ambiguity and Openness, were related to effective job performance (Kilcullen, White, Zacarro, & Parker, 2000). Social Perceptiveness and Interpersonal Skills were most important to supervisory performance at lower levels. Tolerance for Ambiguity and Openness were stronger determinants of successful leadership at higher levels of responsibility where the nature of the work was less structured and ill-defined.

In developing the BIQ, all candidate items were reviewed for construct relevance, response variability, readability, non-intrusiveness, and neutrality with respect to social desirability. The surviving items were pilot tested and revised based on internal consistency reliability and susceptibility to faking.

Response Formats and Scoring

Soldiers were asked to indicate the extent to which each of the 156 BIQ items described themselves using a four- to five-option Likert rating scale. Response options on the BIQ were scored rationally, based on the presumed relationship of the item responses to the underlying psychological construct. Scores for each BIQ scale were calculated by averaging Soldiers' responses across items corresponding to the construct reflected by the given BIQ scale.

Results

Data Preparation

Soldiers' responses to BIQ items were screened prior to conducting any validation analyses. BIQ data were first reviewed for missing responses. We retained only individuals who responded to at least 90% of BIQ items (141 out of 156). Out of the 1,877 Soldiers who completed the BIQ, only 37 provided fewer than 141 responses. These 37 Soldiers were eliminated from all further BIQ analyses. Based on a careful review by two psychologists for

evidence of patterned responding (e.g., a Soldier always choosing response option "a"), 26 Soldiers' BIQ data were removed from further analyses. In sum, 1,817 Soldiers had usable BIQ data for subsequent analyses.[31]

Descriptive Statistics

Table 8.18 presents coefficients alpha and intercorrelations for the BIQ scores by pay grade. With the exception of coefficients alpha for the BIQ Tolerance for Ambiguity and BIQ Interpersonal Skill scales, all BIQ scales had coefficients alpha greater than .60.

Descriptive statistics for the eight BIQ scores are presented by subgroup in Tables 8.18 through 8.34. Raw and conditional effect sizes were calculated by methods summarized in Chapter 3. Overall, Tables 8.18 through 8.34 provide little evidence for large subgroup differences on any of the BIQ scores.

Validity Estimates

Zero-order correlations were computed seperately for E5 and E6 Soldiers, and differences between corresponding correlations (across pay grades) were tested for statistical significance. All correlations were corrected for criterion unreliability and direct range restriction on the predictor using methods described earlier (cf. Chapter 3). Raw and corrected correlations are presented in Table 8.35.

The Leadership and Social Perceptiveness scores exhibited moderate to high validity estimates against observed performance (.33 and .21, respectively) and expected future performance (.42 and .25, respectively) among E5 Soldiers, but significantly lower estimates E6 Soldiers (.05 and -.02 for observed performance, respectively; .09 and .04 for expected performance, respectively). These observed differences between E5 and E6 validity estimates *were* statistically significant ($p < .05$). Like the pay grade differences found for several of the AIM scores, the differences in E5 and E6 validity estimates may indicate these temperament constructs are less predictive of successful performance at the E6 level than at the E5 level, or simply that E6 Soldiers have the requisite levels of the these temperaments.

Several of the BIQ scores (Hostility to Authority, Manipulativeness, and Interpersonal Skill) showed low but statistically significant validity estimates against observed and expected future performance for E5 and E6 Soldiers. Validity estimates for Tolerance for Ambiguity were low for both E5 and E6 Soldiers, yet these estimates were statistically significant for E5 Soldiers. The differences between E5 and E6 correlations for this set of BIQ scores were neither sizable nor statistically significant. Lastly, the BIQ Social Maturity and Openness scores exhibited little to no criterion-related validity for E5 and E6 Soldiers.

[31] Some soldiers who had more than 10% of their BIQ responses missing also exhibited patterned responding. Thus, the reported number of soldiers eliminated for missing data and the reported number of soldiers eliminated for patterned responding are not mutually exclusive.

Table 8.18. BIQ Score Intercorrelations and Reliability Estimates

Predictor	BIQ Host Auth	BIQ Manip	BIQ Social Percept	BIQ Social Mat	BIQ Toler Ambig	BIQ Open	BIQ Leader	BIQ Interpers Skill
E4 Soldiers								
BIQ Hostility to Authority	(.71)							
BIQ Manipulativeness	.61*	(.75)						
BIQ Social Perceptiveness	.04	-.14*	(.80)					
BIQ Social Maturity	-.67*	-.66*	-.02	(.74)				
BIQ Tolerance for Ambiguity	-.29*	-.36*	.23*	.24*	(.41)			
BIQ Openness	.04	-.06	.46*	-.03	.33*	(.66)		
BIQ Leadership	.04	-.09*	.66*	-.06	.27*	.42*	(.79)	
BIQ Interpersonal Skill	-.54*	-.52*	.24*	.41*	.36*	.08*	.20*	(.52)
E5 Soldiers								
BIQ Hostility to Authority	(.72)							
BIQ Manipulativeness	.59*	(.77)						
BIQ Social Perceptiveness	.09*	-.15*	(.83)					
BIQ Social Maturity	-.62*	-.59*	-.08*	(.69)				
BIQ Tolerance for Ambiguity	-.35*	-.39*	.24*	.22*	(.52)			
BIQ Openness	.13*	.01	.48*	-.11*	.27*	(.68)		
BIQ Leadership	.04	-.18*	.69*	-.06*	.31*	.45*	(.82)	
BIQ Interpersonal Skill	-.53*	-.55*	.20*	.44*	.40*	.12*	.19*	(.52)
E6 Soldiers								
BIQ Hostility to Authority	(.71)							
BIQ Manipulativeness	.57*	(.75)						
BIQ Social Perceptiveness	.07*	-.12*	(.83)					
BIQ Social Maturity	-.55*	-.61*	-.13*	(.67)				
BIQ Tolerance for Ambiguity	-.25*	-.34*	.24*	.18*	(.34)			
BIQ Openness	.13*	.00	.44*	-.13*	.18*	(.62)		
BIQ Leadership	.02	-.21*	.61*	-.09*	.25*	.44*	(.80)	
BIQ Interpersonal Skill	-.54*	-.55*	.15*	.43*	.27*	.03	.12*	(.56)

Note. n_{E4} = 430; n_{E5} = 862; n_{E6} = 522-523. Correlations are uncorrected. Internal consistency reliability estimates (coefficients alpha) are in parentheses.

*$p < .05$ (one-tailed).

Table 8.19. Subgroup Differences by Pay Grade, Gender, and Race for BIQ Hostility to Authority

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	2.84	0.55	-0.39	.002	65	2.95	0.56	-0.26	.209
Male	351	3.07	0.57			307	3.09	0.55		
Race										
Black	87	3.06	0.60	0.07	.556	84	3.09	0.56	0.25	.132
White	290	3.02	0.55			288	2.95	0.55		
E5										
Gender										
Female	109	2.84	0.63	-0.05	.649	90	2.91	0.64	0.08	.127
Male	749	2.86	0.58			655	2.86	0.58		
Race										
Black	238	2.86	0.63	0.04	.658	237	2.93	0.62	0.17	.180
White	509	2.84	0.57			508	2.84	0.57		
E6										
Gender										
Female	53	2.67	0.48	-0.09	.527	42	2.79	0.54	0.18	.401
Male	469	2.72	0.55			409	2.70	0.55		
Race										
Black	172	2.71	0.54	-0.01	.905	171	2.76	0.55	0.06	.652
White	283	2.72	0.55			280	2.73	0.55		
Grade										
E5	862	2.86	0.59	-0.29	<.001	745	2.89	0.59	-0.24	.044
E4	430	3.03	0.57			372	3.02	0.55		
E6	523	2.72	0.55	-0.24	<.001	451	2.74	0.55	-0.24	.025
E5	862	2.86	0.59			745	2.89	0.59		
E6	523	2.72	0.55	-0.54	<.001	451	2.74	0.55	-0.50	<.001
E4	430	3.03	0.57			372	3.02	0.55		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means

Table 8.20. Differences between CMF Clusters for BIQ Hostility to Authority

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	66	54	2.83	2.88	0.57	0.54	—	.	0.51	0.17	.	0.57
2. INT	20	19	2.76	2.82	0.63	0.58	-0.13	—	.	.	.	.
3. CBO	171	156	3.12	3.16	0.56	0.56	0.51**	0.64*	—	-0.34	.	0.07
4. LOG	134	112	3.00	2.97	0.56	0.54	0.30*	0.43	-0.21	—	.	0.41
5. CPA	12	10	3.04	3.10	0.43	0.49	.	.	.	.	—	.
6. COM	24	21	3.19	3.19	0.64	0.46	0.62*	0.75*	0.11	0.32	.	—
Overall	430		3.03		0.57							
E5 Soldiers												
1. ADM	82	67	2.79	2.73	0.55	0.51	—	0.22	0.46	0.18	0.23	0.50
2. INT	37	35	2.79	2.86	0.56	0.55	0.02	—	0.24	-0.04	0.02	0.28 *
3. CBO	325	280	2.91	3.00	0.59	0.58	0.21	0.20	—	-0.28	-0.23	0.04
4. LOG	280	240	2.84	2.84	0.60	0.61	0.09	0.07	-0.13	—	0.05	0.32
5. CPA	78	72	2.81	2.87	0.62	0.64	0.04	0.02	-0.18	-0.05	—	0.27
6. COM	58	51	2.91	3.03	0.52	0.54	0.21*	0.19	-0.01	0.12	0.17	—
Overall	862		2.86		0.59							
E6 Soldiers												
1. ADM	53	46	2.65	2.63	0.54	0.56	—	.	0.48	0.19	0.13	0.18
2. INT	21	16	2.64	2.79	0.46	0.45	-0.02	—	.	.	.	.
3. CBO	200	174	2.79	2.89	0.55	0.55	0.25	0.27	—	-0.29	-0.35	-0.30
4. LOG	155	135	2.70	2.73	0.52	0.51	0.09	0.11	-0.16	—	-0.06	-0.01
5. CPA	63	54	2.68	2.70	0.59	0.61	0.06	0.08	-0.19	-0.03	—	0.05
6. COM	31	26	2.55	2.73	0.62	0.67	-0.18	-0.17	-0.43*	-0.27	-0.24	—
Overall	523		2.72		0.55							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.21. Subgroup Differences by Pay Grade, Gender, and Race for BIQ Manipulativeness

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	2.41	0.44	-0.35	.005	65	2.41	0.48	-0.35	.754
Male	351	2.61	0.57			307	2.61	0.55		
Race										
Black	87	2.66	0.55	0.19	.124	84	2.57	0.54	0.22	.149
White	290	2.56	0.56			288	2.45	0.54		
E5										
Gender										
Female	109	2.34	0.61	-0.13	.210	90	2.31	0.58	-0.21	.258
Male	749	2.41	0.54			655	2.42	0.54		
Race										
Black	238	2.46	0.59	0.18	.026	237	2.39	0.57	0.12	.316
White	509	2.37	0.53			508	2.33	0.53		
E6										
Gender										
Female	53	2.19	0.41	-0.22	.126	42	2.22	0.41	-0.13	.577
Male	469	2.30	0.49			409	2.28	0.48		
Race										
Black	172	2.33	0.50	0.13	.213	171	2.26	0.50	0.02	.887
White	283	2.27	0.45			280	2.25	0.45		
Grade										
E5	862	2.40	0.55	-0.30	<.001	745	2.36	0.54	-0.28	.015
E4	430	2.57	0.56			372	2.51	0.54		
E6	523	2.29	0.49	-0.20	<.001	451	2.25	0.47	-0.20	.066
E5	862	2.40	0.55			745	2.36	0.54		
E6	523	2.29	0.49	-0.50	<.001	451	2.25	0.47	-0.47	<.001
E4	430	2.57	0.56			372	2.51	0.54		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.22. Differences between CMF Clusters for BIQ Manipulativeness

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	66	54	2.45	2.50	0.53	0.52	—	.	0.14	0.06	.	-0.02
2. INT	20	19	2.30	2.41	0.43	0.43	-0.27	—	.	.	.	.
3. CBO	171	156	2.70	2.58	0.60	0.60	0.45**	0.73**	—	-0.09	.	-0.16
4. LOG	134	112	2.54	2.53	0.53	0.52	0.16	0.43*	-0.30*	—	.	-0.08
5. CPA	12	10	2.48	2.57	0.49	0.44	.	.	.	.	—	.
6. COM	24	21	2.42	2.49	0.40	0.31	-0.06	0.22	-0.51*	-0.21	.	—
Overall	430		2.57		0.56							
E5 Soldiers												
1. ADM	82	67	2.45	2.43	0.58	0.58	—	-0.22	-0.23	-0.08	-0.11	-0.14
2. INT	37	35	2.23	2.31	0.55	0.56	-0.40	—	-0.01	0.14	0.11	0.08
3. CBO	325	280	2.42	2.31	0.57	0.57	-0.05	0.35*	—	0.16	0.12	0.09
4. LOG	280	240	2.41	2.39	0.53	0.52	-0.06	0.34	-0.01	—	-0.03	-0.06
5. CPA	78	72	2.40	2.37	0.54	0.53	-0.09	0.31	-0.04	-0.03	—	-0.03
6. COM	58	51	2.31	2.36	0.46	0.44	-0.25	0.15	-0.20	-0.19	-0.16	—
Overall	862		2.40		0.55							
E6 Soldiers												
1. ADM	53	46	2.27	2.29	0.41	0.40	—	.	-0.08	-0.10	-0.25	-0.06
2. INT	21	16	2.23	2.29	0.45	0.45	-0.08	—	.	.	.	.
3. CBO	200	174	2.35	2.25	0.50	0.48	0.18	0.26	—	-0.02	-0.16	0.02
4. LOG	155	135	2.27	2.25	0.49	0.47	0.01	0.09	-0.17	—	-0.15	0.04
5. CPA	63	54	2.25	2.18	0.50	0.49	-0.04	0.04	-0.22	-0.05	—	0.19
6. COM	31	26	2.14	2.26	0.46	0.49	-0.26	-0.18	-0.44*	-0.27	-0.22	—
Overall	523		2.29		0.49							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.23. Subgroup Differences by Pay Grade, Gender, and Race for BIQ Social Perceptiveness

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	3.61	0.48	0.16	.217	65	3.72	0.5	0.15	.474
Male	351	3.53	0.53			307	3.64	0.5		
Race										
Black	87	3.57	0.52	0.05	.658	84	3.69	0.52	0.02	.880
White	290	3.55	0.5			288	3.67	0.5		
E5										
Gender										
Female	109	3.54	0.5	-0.02	.858	90	3.65	0.51	0.19	.279
Male	749	3.55	0.53			655	3.55	0.53		
Race										
Black	238	3.52	0.55	-0.09	.249	237	3.57	0.55	-0.13	.283
White	509	3.57	0.52			508	3.63	0.52		
E6										
Gender										
Female	53	3.42	0.49	-0.27	.064	42	3.49	0.5	-0.19	.382
Male	469	3.55	0.48			409	3.58	0.48		
Race										
Black	172	3.59	0.5	0.10	.327	171	3.55	0.49	0.07	.647
White	283	3.54	0.48			280	3.52	0.48		
Grade										
E5	862	3.54	0.53	0.00	.998	745	3.6	0.53	-0.16	.185
E4	430	3.54	0.53			372	3.68	0.5		
E6	523	3.54	0.49	-0.01	.884	451	3.53	0.48	-0.13	.242
E5	862	3.54	0.53			745	3.6	0.53		
E6	523	3.54	0.49	-0.01	.885	451	3.53	0.48	-0.29	.031
E4	430	3.54	0.53			372	3.68	0.5		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.24. Differences between CMF Clusters for BIQ Social Perceptiveness

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	66	54	3.6	3.64	0.52	0.50	—	.	0.28	-0.07	.	0.04
2. INT	20	19	3.81	3.63	0.60	0.56	0.40	—	.	.	.	.
3. CBO	171	156	3.49	3.78	0.48	0.48	-0.21	-0.61*	—	-0.35	.	-0.24
4. LOG	134	112	3.51	3.60	0.53	0.50	-0.18	-0.58*	0.03	—	.	0.11
5. CPA	12	10	3.64	3.78	0.59	0.47	.	.	.	.	—	.
6. COM	24	21	3.71	3.66	0.64	0.63	0.20	-0.19	0.41*	0.38	.	—
Overall	430		3.54		0.53							
E5 Soldiers												
1. ADM	82	67	3.51	3.54	0.51	0.50	—	0.19	0.63	0.01	0.05	-0.14
2. INT	37	35	3.74	3.64	0.41	0.39	0.43*	—	0.44	-0.18	-0.14	-0.33
3. CBO	325	280	3.58	3.87	0.54	0.54	0.12	-0.30*	—	-0.61	-0.58	-0.76
4. LOG	280	240	3.50	3.54	0.56	0.57	-0.02	-0.44**	-0.14	—	0.04	-0.15
5. CPA	78	72	3.54	3.56	0.48	0.50	0.05	-0.38*	-0.07	0.07	—	-0.19
6. COM	58	51	3.48	3.46	0.42	0.41	-0.07	-0.50**	-0.19	-0.05	-0.12	—
Overall	862		3.54		0.53							
E6 Soldiers												
1. ADM	53	46	3.50	3.50	0.53	0.50	—	.	0.61	0.07	-0.13	0.16
2. INT	21	16	3.51	3.37	0.58	0.54	0.02	—	.	.	.	.
3. CBO	200	174	3.56	3.79	0.46	0.47	0.13	0.11	—	-0.53	-0.73	-0.45
4. LOG	155	135	3.53	3.53	0.46	0.47	0.06	0.04	-0.07	—	-0.20	0.09
5. CPA	63	54	3.48	3.44	0.47	0.45	-0.05	-0.07	-0.18	-0.11	—	0.29
6. COM	31	26	3.63	3.58	0.63	0.63	0.26	0.24	0.13	0.20	0.31	—
Overall	523		3.54		0.49							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.25. Subgroup Differences by Pay Grade, Gender, and Race for BIQ Social Maturity

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	3.57	0.53	0.79	<.001	65	3.54	0.55	0.72	<.001
Male	351	3.06	0.64			307	3.08	0.64		
Race										
Black	87	3.21	0.72	0.15	.223	84	3.30	0.65	-0.04	.800
White	290	3.12	0.64			288	3.32	0.62		
E5										
Gender										
Female	109	3.62	0.55	0.52	<.001	90	3.63	0.57	0.51	.006
Male	749	3.32	0.59			655	3.33	0.58		
Race										
Black	238	3.43	0.61	0.16	.042	237	3.51	0.59	0.09	.458
White	509	3.33	0.58			508	3.45	0.57		
E6										
Gender										
Female	53	3.75	0.42	0.51	<.001	42	3.77	0.45	0.48	.034
Male	469	3.48	0.53			409	3.52	0.52		
Race										
Black	172	3.59	0.52	0.25	.010	171	3.71	0.52	0.22	.154
White	283	3.46	0.52			280	3.59	0.52		
Grade										
E5	862	3.35	0.59	0.30	<.001	745	3.48	0.58	0.27	.011
E4	430	3.15	0.65			372	3.31	0.63		
E6	523	3.50	0.53	0.26	<.001	451	3.65	0.52	0.29	.009
E5	862	3.35	0.59			745	3.48	0.58		
E6	523	3.50	0.53	0.54	<.001	451	3.65	0.52	0.54	<.001
E4	430	3.15	0.65			372	3.31	0.63		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.26. Differences between CMF Clusters for BIQ Social Maturity

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	66	54	3.44	3.42	0.58	0.55	—	.	-0.28	-0.20	.	-0.25
2. INT	20	19	3.46	3.52	0.53	0.53	0.03	—	.	.	.	.
3. CBO	171	156	2.95	3.24	0.65	0.66	-0.75**	-0.78**	—	0.08	.	0.02
4. LOG	134	112	3.21	3.29	0.64	0.62	-0.36**	-0.39*	0.39**	—	.	-0.06
5. CPA	12	10	3.2	3.13	0.54	0.52	.	.	.	.	—	.
6. COM	24	21	3.2	3.26	0.73	0.7	-0.37	-0.40	0.38	-0.01	.	—
Overall	430		3.15		0.65							
E5 Soldiers												
1. ADM	82	67	3.42	3.49	0.58	0.54	—	0.02	0.06	-0.01	-0.17	-0.00
2. INT	37	35	3.38	3.5	0.51	0.51	-0.07	—	0.04	-0.03	-0.20	-0.03
3. CBO	325	280	3.27	3.53	0.61	0.59	-0.26*	-0.19	—	-0.07	-0.24	-0.07
4. LOG	280	240	3.42	3.48	0.61	0.6	-0.00	0.07	0.26*	—	-0.16	0.01
5. CPA	78	72	3.3	3.39	0.53	0.52	-0.21	-0.14	0.05	-0.21	—	0.17
6. COM	58	51	3.46	3.49	0.52	0.54	0.06	0.13	0.32*	0.06	0.27	—
Overall	862		3.35		0.59							
E6 Soldiers												
1. ADM	53	46	3.62	3.64	0.43	0.45	—	.	-0.03	-0.03	-0.05	0.16
2. INT	21	16	3.5	3.67	0.44	0.5	-0.22	—	.	.	.	.
3. CBO	200	174	3.42	3.62	0.54	0.54	-0.39**	-0.16	—	-0.00	-0.02	0.19
4. LOG	155	135	3.55	3.62	0.54	0.52	-0.12	0.10	0.26*	—	-0.02	0.19
5. CPA	63	54	3.45	3.61	0.52	0.49	-0.32*	-0.09	0.07	-0.19	—	0.21
6. COM	31	26	3.72	3.72	0.45	0.5	0.18	0.41	0.57**	0.31	0.50*	—
Overall	523		3.5		0.53							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.27. Subgroup Differences by Pay Grade, Gender, and Race for BIQ Tolerance for Ambiguity

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	3.18	0.39	0.10	.445	65	3.21	0.39	0.01	.529
Male	351	3.14	0.40			307	3.20	0.40		
Race										
Black	87	3.10	0.36	-0.16	.174	84	3.18	0.38	-0.13	.245
White	290	3.17	0.41			288	3.23	0.40		
E5										
Gender										
Female	109	3.15	0.35	-0.16	.111	90	3.16	0.37	-0.11	.611
Male	749	3.22	0.43			655	3.21	0.41		
Race										
Black	238	3.12	0.40	-0.39	<.001	237	3.11	0.41	-0.35	.277
White	509	3.28	0.41			508	3.26	0.41		
E6										
Gender										
Female	53	3.23	0.28	0.04	.804	42	3.26	0.27	0.03	.521
Male	469	3.22	0.37			409	3.25	0.36		
Race										
Black	172	3.15	0.33	-0.32	<.001	171	3.21	0.33	-0.22	.006
White	283	3.27	0.37			280	3.29	0.37		
Grade										
E5	862	3.21	0.42	0.16	<.001	745	3.18	0.41	-0.05	.661
E4	430	3.15	0.40			372	3.20	0.40		
E6	523	3.22	0.36	0.02	.735	451	3.25	0.36	0.17	.114
E5	862	3.21	0.42			745	3.18	0.41		
E6	523	3.22	0.36	0.19	.002	451	3.25	0.36	0.12	.344
E4	430	3.15	0.40			372	3.20	0.40		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.28. Differences between CMF Clusters for BIQ Tolerance for Ambiguity

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	66	54	3.17	3.17	0.34	0.34	—	.	-0.22	-0.07	.	-0.14
2. INT	20	19	3.36	3.25	0.45	0.46	0.46*	—	.	.	.	.
3. CBO	171	156	3.07	3.09	0.45	0.45	-0.25*	-0.71**	—	0.15	.	0.08
4. LOG	134	112	3.15	3.15	0.34	0.35	-0.05	-0.51	0.20	—	.	-0.07
5. CPA	12	10	3.44	3.45	0.36	0.30	.	.	.	.	—	.
6. COM	24	21	3.20	3.12	0.31	0.34	0.08	-0.38	0.33	0.12	.	—
Overall	430		3.15		0.40							
E5 Soldiers												
1. ADM	82	67	3.18	3.20	0.43	0.44	—	-0.16	-0.06	-0.01	0.28	-0.25
2. INT	37	35	3.30	3.13	0.46	0.42	0.28	—	0.10	0.16	0.44	-0.09
3. CBO	325	280	3.22	3.17	0.44	0.41	0.08	-0.20	—	0.06	0.34	-0.19
4. LOG	280	240	3.19	3.20	0.41	0.40	0.02	-0.26	-0.06	—	0.29	-0.25
5. CPA	78	72	3.28	3.31	0.40	0.39	0.23	-0.05	0.16	0.22	—	-0.53
6. COM	58	51	3.18	3.09	0.43	0.39	-0.01	-0.29	-0.08	-0.02	-0.24 *	—
Overall	862		3.21		0.42							
E6 Soldiers												
1. ADM	53	46	3.21	3.21	0.35	0.36	—	.	0.05	-0.10	0.43	0.24
2. INT	21	16	3.36	3.26	0.37	0.33	0.41	—	.	.	.	.
3. CBO	200	174	3.20	3.23	0.37	0.37	-0.02	-0.43	—	-0.15	0.38	0.19
4. LOG	155	135	3.15	3.17	0.34	0.35	-0.17	-0.57*	-0.15	—	0.52 *	0.34
5. CPA	63	54	3.33	3.36	0.36	0.35	0.34	-0.07	0.36*	0.50**	—	-0.18
6. COM	31	26	3.37	3.29	0.38	0.38	0.45*	0.04	0.47*	0.61**	0.11	—
Overall	523		3.22		0.36							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.29. Subgroup Differences by Pay Grade, Gender, and Race for BIQ Openness

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	3.41	0.49	-0.05	.671	65	3.25	0.50	-0.52	.012
Male	351	3.43	0.50			307	3.50	0.49		
Race										
Black	87	3.49	0.49	0.16	.178	84	3.39	0.51	0.04	.783
White	290	3.41	0.49			288	3.37	0.49		
E5										
Gender										
Female	109	3.33	0.45	-0.11	.260	90	3.21	0.44	-0.41	.022
Male	749	3.38	0.51			655	3.41	0.50		
Race										
Black	238	3.39	0.45	0.04	.557	237	3.27	0.44	-0.16	.158
White	509	3.37	0.52			508	3.35	0.52		
E6										
Gender										
Female	53	3.21	0.47	-0.27	.068	42	3.10	0.48	-0.63	.005
Male	469	3.33	0.45			409	3.38	0.44		
Race										
Black	172	3.36	0.43	0.16	.084	171	3.23	0.42	-0.05	.746
White	283	3.28	0.47			280	3.25	0.46		
Grade										
E5	862	3.37	0.51	-0.12	.017	745	3.31	0.50	-0.14	.221
E4	430	3.43	0.50			372	3.38	0.49		
E6	523	3.32	0.45	-0.11	.082	451	3.24	0.45	-0.13	.213
E5	862	3.37	0.51			745	3.31	0.50		
E6	523	3.32	0.45	-0.22	<.001	451	3.24	0.45	-0.28	.034
E4	430	3.43	0.50			372	3.38	0.49		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.30. Differences between CMF Clusters for BIQ Openness

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	66	54	3.49	3.48	0.51	0.52	—	.	-0.63	-0.07	.	-0.18
2. INT	20	19	3.62	3.37	0.52	0.42	0.25	—	.	.	.	.
3. CBO	171	156	3.42	3.17	0.49	0.49	-0.14	-0.39	—	0.56	.	0.44
4. LOG	134	112	3.39	3.44	0.50	0.49	-0.20	-0.46	-0.07	—	.	-0.11
5. CPA	12	10	3.27	3.42	0.42	0.44	.	.	.	.	—	.
6. COM	24	21	3.45	3.39	0.48	0.49	-0.08	-0.33	0.06	0.13	.	—
Overall	430		3.43		0.50							
E5 Soldiers												
1. ADM	82	67	3.35	3.40	0.59	0.53	—	-0.16	-0.57	-0.06	-0.12	-0.24
2. INT	37	35	3.57	3.32	0.51	0.46	0.43	—	-0.41	0.11	0.04	-0.08
3. CBO	325	280	3.40	3.12	0.51	0.50	0.11	-0.33	—	0.51	0.45	0.32
4. LOG	280	240	3.34	3.38	0.49	0.49	-0.01	-0.45*	-0.12	—	-0.07	-0.19
5. CPA	78	72	3.32	3.34	0.51	0.53	-0.06	-0.49*	-0.16	-0.04	—	-0.12
6. COM	58	51	3.36	3.28	0.41	0.42	0.03	-0.41	-0.08	0.04	0.08	—
Overall	862		3.37		0.51							
E6 Soldiers												
1. ADM	53	46	3.31	3.28	0.43	0.44	—	.	-0.57	0.01	0.09	0.08
2. INT	21	16	3.37	3.20	0.44	0.47	0.14	—	.	.	.	.
3. CBO	200	174	3.34	3.03	0.45	0.45	0.06	-0.08	—	0.58	0.66	0.64
4. LOG	155	135	3.30	3.29	0.45	0.44	-0.02	-0.16	-0.07	—	0.08	0.07
5. CPA	63	54	3.29	3.32	0.52	0.50	-0.06	-0.20	-0.11	-0.04	—	-0.01
6. COM	31	26	3.37	3.32	0.41	0.38	0.13	-0.01	0.07	0.14	0.18	—
Overall	523		3.32		0.45							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.31. Subgroup Differences by Pay Grade, Gender, and Race BIQ Leadership

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	3.37	0.52	-0.01	.951	65	3.46	0.51	0.02	.925
Male	351	3.37	0.50			307	3.45	0.49		
Race										
Black	87	3.42	0.47	0.13	.269	84	3.52	0.47	0.25	.110
White	290	3.36	0.50			288	3.39	0.49		
E5										
Gender										
Female	109	3.50	0.47	0.00	.982	90	3.53	0.46	0.10	.560
Male	749	3.50	0.51			655	3.48	0.51		
Race										
Black	238	3.51	0.50	0.05	.485	237	3.53	0.50	0.12	.300
White	509	3.48	0.51			508	3.47	0.51		
E6										
Gender										
Female	53	3.47	0.50	-0.27	.070	42	3.51	0.55	-0.24	.282
Male	469	3.59	0.44			409	3.61	0.44		
Race										
Black	172	3.57	0.48	-0.02	.814	171	3.59	0.49	0.14	.391
White	283	3.58	0.43			280	3.53	0.43		
Grade										
E5	862	3.50	0.51	0.26	.001	745	3.50	0.50	0.10	.412
E4	430	3.37	0.50			372	3.46	0.49		
E6	523	3.58	0.45	0.15	.011	451	3.56	0.45	0.11	.294
E5	862	3.50	0.51			745	3.50	0.50		
E6	523	3.58	0.45	0.41	<.001	451	3.56	0.45	0.21	.108
E4	430	3.37	0.50			372	3.46	0.49		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.32. Differences between CMF Clusters for BIQ Leadership

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	66	54	3.42	3.43	0.52	0.53	—	.	0.03	-0.15	.	-0.31
2. INT	20	19	3.54	3.51	0.43	0.41	0.23	—	.	.	.	.
3. CBO	171	156	3.37	3.44	0.49	0.49	-0.10	-0.33	—	-0.18	.	-0.34
4. LOG	134	112	3.31	3.35	0.49	0.44	-0.22	-0.45*	-0.12	—	.	-0.16
5. CPA	12	10	3.60	3.74	0.63	0.59	.	.	.	.	—	.
6. COM	24	21	3.36	3.27	0.58	0.62	-0.11	-0.35	-0.01	0.10	.	—
Overall	430		3.37		0.50							
E5 Soldiers												
1. ADM	82	67	3.51	3.52	0.53	0.50	—	-0.02	0.19	-0.14	0.10	-0.38
2. INT	37	35	3.52	3.51	0.50	0.52	0.02	—	0.21	-0.11	0.12	-0.36
3. CBO	325	280	3.55	3.62	0.51	0.51	0.09	0.06	—	-0.33	-0.09	-0.58
4. LOG	280	240	3.47	3.46	0.52	0.51	-0.08	-0.10	-0.17*	—	0.23	-0.25
5. CPA	78	72	3.45	3.57	0.44	0.46	-0.11	-0.13	-0.20	-0.03	—	-0.48 *
6. COM	58	51	3.38	3.33	0.44	0.46	-0.26	-0.28	-0.35*	-0.18	-0.15 *	—
Overall	862		3.50		0.51							
E6 Soldiers												
1. ADM	53	46	3.55	3.54	0.49	0.44	—	.	0.15	-0.17	0.22	-0.09
2. INT	21	16	3.69	3.61	0.43	0.42	0.31	—	.	.	.	.
3. CBO	200	174	3.60	3.61	0.45	0.44	0.11	-0.20	—	-0.32	0.07	-0.24
4. LOG	155	135	3.54	3.46	0.44	0.46	-0.01	-0.33	-0.12	—	0.39	0.08
5. CPA	63	54	3.58	3.64	0.41	0.44	0.07	-0.25	-0.04	0.08	—	-0.31
6. COM	31	26	3.59	3.50	0.56	0.56	0.10	-0.21	-0.01	0.12	0.03	—
Overall	523		3.58		0.45							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.33. Subgroup Differences by Pay Grade, Gender, and Race for BIQ Interpersonal Skill

	Raw					Conditional				
Group	*n*	*M*	*SD*	Effect Size	*p*	*n*	*M*	*SD*	Effect Size	*p*
E4										
Gender										
Female	74	3.17	0.40	0.22	.076	65	3.20	0.39	0.34	.086
Male	351	3.06	0.45			307	3.05	0.45		
Race										
Black	87	2.91	0.44	-0.50	<.001	84	2.99	0.44	-0.61	<.001
White	290	3.13	0.44			288	3.26	0.44		
E5										
Gender										
Female	109	3.09	0.46	-0.17	.108	90	3.16	0.47	-0.07	.720
Male	749	3.16	0.42			655	3.19	0.42		
Race										
Black	238	3.11	0.44	-0.15	.054	237	3.12	0.44	-0.26	.040
White	509	3.18	0.43			508	3.23	0.42		
E6										
Gender										
Female	53	3.21	0.37	0.01	.958	42	3.28	0.36	0.13	.541
Male	468	3.21	0.43			408	3.22	0.43		
Race										
Black	171	3.19	0.43	-0.05	.606	170	3.22	0.43	-0.14	.343
White	283	3.21	0.42			280	3.28	0.42		
Grade										
E5	862	3.15	0.43	0.15	.002	745	3.17	0.43	0.11	.334
E4	430	3.08	0.45			372	3.12	0.44		
E6	522	3.21	0.42	0.13	.029	450	3.25	0.42	0.19	.094
E5	862	3.15	0.43			745	3.17	0.43		
E6	522	3.21	0.42	0.28	<.001	450	3.25	0.42	0.29	.024
E4	430	3.08	0.45			372	3.12	0.44		

Note. Raw effect sizes calculated as (*M* of non-referent group – *M* of referent group)/*SD* referent group. Referent groups (e.g., males) are listed second in each pair. *p*-values reflect significance levels for two-tailed *t*-tests of differences between subgroup means.

Table 8.34. Differences between CMF Clusters for BIQ Interpersonal Skill

CMFC	*n*		*M*		*SD*		Effect Size					
	Raw	Con	Raw	Con	Raw	Con	1. ADM	2. INT	3. CBO	4. LOG	5. CPA	6. COM
E4 Soldiers												
1. ADM	66	54	3.14	3.12	0.46	0.47	—	.	0.25	-0.18	.	-0.12
2. INT	20	19	3.37	3.18	0.44	0.46	0.53*	—	.	.	.	.
3. CBO	171	156	3.03	3.23	0.44	0.44	-0.24	-0.77**	—	-0.43	.	-0.37
4. LOG	134	112	3.06	3.04	0.43	0.43	-0.19	-0.71**	0.06	—	.	0.06
5. CPA	12	10	3.22	3.11	0.40	0.36	.	.	.	.	—	.
6. COM	24	21	3.17	3.06	0.43	0.38	0.06	-0.46	0.31	0.25	.	—
Overall	430		3.08		0.45							
E5 Soldiers												
1. ADM	82	67	3.25	3.28	0.46	0.43	—	-0.30	0.06	-0.48**	-0.36	-0.45 *
2. INT	37	35	3.28	3.15	0.45	0.47	0.08	—	0.37	-0.18	-0.05	-0.15
3. CBO	325	280	3.14	3.31	0.44	0.44	-0.25	-0.33	—	-0.54	-0.42	-0.52
4. LOG	280	240	3.11	3.07	0.42	0.42	-0.31 *	-0.39*	-0.06	—	0.12	0.02
5. CPA	78	72	3.17	3.13	0.41	0.41	-0.19	-0.26	0.07	0.13	—	-0.10
6. COM	58	51	3.16	3.09	0.40	0.37	-0.20	-0.27	0.06	0.11	-0.01	—
Overall	862		3.15		0.43							
E6 Soldiers												
1. ADM	52	45	3.22	3.22	0.40	0.43	—	.	0.36	-0.03	0.18	0.24
2. INT	21	16	3.21	3.09	0.51	0.49	-0.01	—	.	.	.	.
3. CBO	200	174	3.15	3.37	0.42	0.41	-0.16	-0.15	—	-0.39	-0.18	-0.12
4. LOG	155	135	3.22	3.21	0.44	0.43	0.00	0.01	0.16	—	0.21	0.27
5. CPA	63	54	3.27	3.30	0.39	0.39	0.12	0.13	0.29*	0.12	—	0.06
6. COM	31	26	3.42	3.32	0.42	0.42	0.47*	0.48	0.64**	0.47*	0.35	—
Overall	522		3.21		0.42							

Note. CMFC = Career Management Field Cluster; ADM = Administration; INT = Intelligence; CBO = Combat Operations; LOG = Logistics; CPA = Civil & Public Affairs; COM = Communications. Raw = Raw statistic; Con = Conditional statistic. Raw effect sizes calculated as (*M* of higher-numbered category – *M* of lower-numbered category)/overall *SD*. Raw effect sizes are below the diagonal; conditional effect sizes are above the diagonal. Conditional effect sizes control for differences due to gender and race.

*$p < .05$. **$p < .01$. All significance tests are two-tailed.

Table 8.35. Corrected and Raw Correlations between BIQ Scores and Criteria for E5 and E6 Soldiers

Criterion	Predictor							
	BIQ Hostility to Authority	BIQ Manipulat	BIQ Social Perceptive	BIQ Social Maturity	BIQ Tolerance Ambiguity	BIQ Openness	BIQ Leadership	BIQ Interpersonal Skill
E5 Soldiers								
Observed Performance Composite	-.08 (-.06)	-.11 (-.08*)	.21 (.15$_a$*)	.09 (.06)	.18 (.14*)	.06 (.05)	.33 (.25$_a$*)	.16 (.11*)
Expected Future Performance Composite	-.11 (-.07*)	-.11 (-.07*)	.25 (.16$_a$*)	.02 (.01)	.19 (.13*)	.10 (.07)	.42 (.27$_a$*)	.15 (.09*)
Senior NCO Potential Rating	-.10 (-.07)	-.07 (-.05)	.15 (.10*)	.02 (.02)	.11 (.08*)	-.01 (-.01)	.32 (.22$_a$*)	.11 (.07*)
Overall Effectiveness Rating	-.02 (-.01)	-.06 (-.04)	.21 (.13*)	-.01 (-.01)	.14 (.10*)	.05 (.03)	.36 (.24$_a$*)	.12 (.08*)
E6 Soldiers								
Observed Performance Composite	-.17 (-.13*)	-.15 (-.10*)	-.02 (-.01)	.08 (.06)	.07 (.04)	-.09 (-.06)	.05 (.04)	.18 (.14*)
Expected Future Performance Composite	-.15 (-.10*)	-.17 (-.10*)	.04 (.03)	.11 (.07)	.14 (.08)	-.08 (-.05)	.09 (.06)	.21 (.14*)
Senior NCO Potential Rating	-.18 (-.12*)	-.13 (-.08)	-.01 (.00)	.01 (.00)	.08 (.05)	-.09 (-.06)	.09 (.06)	.11 (.08)
Overall Effectiveness Rating	-.13 (-.09*)	-.03 (-.02)	.10 (.06)	-.07 (-.04)	.07 (.04)	-.04 (-.03)	.11 (.07)	.09 (.06)

Note. n_{E5} = 592-597; n_{E6} = 369-376. Correlations corrected for criterion unreliability and for direct range restriction on the predictor appear outside of the parentheses. Raw correlations appear inside parentheses. The "a" subscripts on E5 correlations indicate that corresponding E5 and E6 correlations were significantly different from each other, $p < .05$ (two-tailed).

* $p < .05$ (one-tailed).

Differential Prediction Analyses

Table 8.36 presents the results of differential prediction analyses for BIQ scores by pay grade and criterion, examining gender and race as the demographic variables of interest. [32] Overall, the results provide little evidence of differential prediction. In only one case was differential prediction evident for race-based comparisons. Specifically, the Manipulativeness score was more predictive of expected future performance for white E5 Soldiers ($b = 0.15$) than for black E5 Soldiers ($b = -0.05$). Differential prediction was evident in only three cases for gender-based comparisons (Openness for E5 Soldiers with both criteria, and Tolerance for Ambiguity for E5 Soldiers with observed performance).

As was the case with the AIM scores, evidence for intercept bias emerged only for gender-based comparisons when predicting expected future performance. Specifically, women scored 0.35 to 0.55 point lower than men on expected performance (at mean levels of the BIQ scores). These findings suggest that the BIQ scores would *overpredict* female Soldiers' expected performance if a common regression equation were used to predict their performance.

BIQ Summary

The BIQ Leadership, Social Perceptiveness, and Tolerance for Ambiguity scores showed promise as predictors for future E4-to-E5 NCO promotion decisions, but not for future E5-to-E6 promotion decisions. The Hostility to Authority, Manipulativeness, and Interpersonal Skill scores showed low but statistically significant validity estimates across pay grades. The Social Maturity and Openness scores showed little evidence of validity.

Although subgroup analyses revealed few differences among subgroups on BIQ scores, there was evidence of intercept bias for gender (females' performance being overpredicted) when predicting expected future performance. Nevertheless, little evidence emerged that suggested BIQ scores (in general) would be differentially predictive of future NCO performance.

[32] All BIQ scores were standardized within pay grade to ease interpretation of the unstandardized regression weights prior to conducting these analyses. The demographic variables were coded as follows for purposes of analysis: race (white = 0, black = 1), gender (male = 0, female = 1).

Table 8.36. Differential Prediction Analyses for BIQ Scores

Criterion/Predictor	Demographic Main Effect		BIQ Score Main Effect				*r*			
			Gender		Race		Gender		Race	
	Gender	Race	M	F	W	B	M	F	W	B
			Observed Performance Composite							
E5 Soldiers										
BIQ Hostility to Authority	-.17	.00	.04	.13	.03	.11	-.04	-.16	-.04	**-.15**
BIQ Manipulativeness	-.21	-.01	.05	.19	.08	.04	-.06	**-.22**	-.09	-.05
BIQ Social Perceptiveness	-.17	-.01	.14	.02	.14	.05	**.17**	.02	**.17**	.07
BIQ Social Maturity	-.21	-.02	.06	.09	.03	.12	.07	.09	.04	**.15**
BIQ Tolerance for Ambiguity	-.12	.04	.10	.34$_a$	.15	.16	**.12**	**.32**	**.16**	**.20**
BIQ Openness	-.17	-.02	.06$_a$	-.21	.02	.02	**.07**	-.19	.02	.02
BIQ Leadership	-.17	-.04	.21	.18	.23	.16	**.26**	.19	**.27**	**.19**
BIQ Interpersonal Skill	-.14	.01	.08	.15	.06	.14	**.09**	.18	.06	**.18**
E6 Soldiers										
BIQ Hostility to Authority	-.17	-.10	.08	.27	.15	.05	**-.11**	**-.30**	**-.21**	-.06
BIQ Manipulativeness	-.22	-.09	.07	.29	.12	.08	**-.10**	-.27	**-.16**	-.09
BIQ Social Perceptiveness	-.20	-.09	-.01	-.08	.01	-.02	-.01	-.10	.01	-.03
BIQ Social Maturity	-.15	-.11	.06	-.07	.10	.02	.08	-.07	**.14**	.02
BIQ Tolerance for Ambiguity	-.17	-.09	.03	.26	.06	.00	.04	.19	.08	.00
BIQ Openness	-.17	-.10	-.07	.09	-.08	.04	-.09	.11	-.11	.05
BIQ Leadership	-.18	-.09	.03	.00	.10	-.02	.04	.00	**.13**	-.02
BIQ Interpersonal Skill	-.18	-.10	.10	.15	.13	.10	**.14**	.17	**.19**	.12
			Expected Future Performance Composite							
E5 Soldiers										
BIQ Hostility to Authority	-.38*	.02	.06	.14	.09	.07	-.06	-.15	**-.09**	-.09
BIQ Manipulativeness	-.41*	.01	.07	.13	.15$_a$	-.05	-.07	-.15	**-.14**	.06
BIQ Social Perceptiveness	-.39*	.02	.16	.09	.18	.07	**.17**	.09	**.19**	.08
BIQ Social Maturity	-.40*	.01	.03	.02	-.01	.01	.03	.02	-.01	.01
BIQ Tolerance for Ambiguity	-.35*	.05	.11	.23	.20	.09	**.11**	**.20**	**.19**	.09
BIQ Openness	-.38*	.00	.09$_a$	-.25	.06	.02	**.10**	**-.22**	.06	.02
BIQ Leadership	-.38*	-.02	.27	.21	.32	.16	**.28**	**.20**	**.33**	**.17**
BIQ Interpersonal Skill	-.36*	.02	.07	.15	.08	.07	.07	.18	.08	.08
E6 Soldiers										
BIQ Hostility to Authority	-.49*	-.13	.07	.33	.12	.09	-.08	-.25	**-.14**	-.09
BIQ Manipulativeness	-.55*	-.12	.09	.33	.09	.16	**-.10**	-.21	-.10	-.14
BIQ Social Perceptiveness	-.55*	-.12	.04	-.24	.04	-.02	.05	-.20	.05	-.02
BIQ Social Maturity	-.51*	-.16	.09	.02	.09	.09	**.10**	.01	.11	.08
BIQ Tolerance for Ambiguity	-.48*	-.10	.06	.59	.09	.07	.07	**.29**	.10	.05
BIQ Openness	-.52*	-.13	-.05	-.07	-.10	.09	-.06	-.06	**-.12**	.07
BIQ Leadership	-.53*	-.12	.07	-.15	.13	-.01	.07	-.13	**.15**	-.01
BIQ Interpersonal Skill	-.51*	-.13	.12	.17	.10	.22	**.14**	.12	**.12**	**.20**

Note. Regression analysis sample sizes: $n_{E5\ Gender}$ = 590-595; $n_{E5\ Race}$ = 510-513; $n_{E6\ Gender}$ = 368-375; $n_{E6\ Race}$ = 323-328. Smaller sample sizes underlie the reported correlations because they were calculated for each subgroup separately. The "a" subscripts on the BIQ main effect values indicate the BIQ-by-demographic interaction term was statistically significant, $p < .05$ (two-tailed). Subscripts are located on the subgroup with the higher value. Correlations are uncorrected. Bolded correlations are statistically significant, $p < .05$ (one-tailed).

*$p < .05$ (two-tailed).

CHAPTER 9: NCO21 PREDICTOR VALIDITY EVIDENCE

Christopher E. Sager, Dan J. Putka, and Gordon A. Waugh
HumRRO

Overview

This chapter addresses issues relevant to the validity of the NCO21 predictor measures. In previous chapters we examined each predictor and criterion measure largely on its own merits. In this chapter we focus on (a) additional evidence for the construct validity of the new predictors developed as part of the NCO21 project[33] and (b) the degree to which additional predictors might improve the predictive validity of the current promotion system. The primary reason for examining these issues is to identify predictor measures that have the potential to improve the future E4-to-E5 and E5-to-E6 Soldier promotion system.

This chapter also examines two other issues that became salient during the course of our analyses: (a) differences in the criterion-related validity of predictors across pay grades and (b) potential differences in the criterion-related validity of predictors across job types (i.e., CMFs). The current Promotion Point Worksheet's scoring and content are the same for promotions to the E5 and E6 pay grades across all MOS. The present findings suggest that it might be useful to establish different standards for promotion to E5 versus promotion to E6, and perhaps for different MOS or MOS groups.

Construct Validity

The goal of construct validation is to support inferences about the meaning of scores from tests that are hypothesized to measure a specified construct. In this case, we want to support (a) using our predictors as measures of the constructs that our job analysis work (Ford et al., 2000) indicated are important KSAs for determining current and future job performance and (b) using our criterion measures as valid measures of current and expected future job performance. Although all of the chapters in this report address the construct validity of the NCO21 measures, this section will focus on the evidence from the (a) relations among the predictor scores and (b) pattern of relations among multiple predictor and performance scores. A subsequent section will address estimation of the criterion-related validity of the predictor measures.

Relations among Predictor Scores

In Chapter 1, Table 1.2 shows the 38 KSAs identified as relevant to the job performance of E5 and E6 Soldiers (Ford et al., 2000), and Table 1.4 indicates which of the eight predictor measures used in this project were designed to assess each KSA. Here we examine the relations among the 26 scores derived from these measures. Table 9.1 shows correlations among these scores for the E5 and E6 Soldier participants.[34] The table orders predictor scores by the

[33] See Chapter 3 for a discussion of the construct validity of, and relations among, the criterion measures.
[34] The correlations among these scores for E4 soldiers are presented in Appendix F. Unless otherwise noted, the E4 sample results were consistent with the description of the E5 sample results.

Table 9.1. Raw Correlations among Predictor Scores by Pay Grade for E5 and E6 Soldiers

Predictors		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	ASVAB GT	.	**.25**	**.09**	.	-.01	-.01	.00	-.04	-.01	**.12**	**-.14**	-.03	.03	**.08**	.05	.04	.06	**.10**	**-.10**	**-.09**	**.13**	.04	**.17**	.00	**.09**	**.09**
2	SJT Composite	**.14**	.	**.19**	.	.03	.01	.00	.06	.01	**.09**	-.03	.01	**.20**	**.14**	**.14**	**.22**	.05	**.19**	**-.22**	**-.20**	.01	**.21**	**.08**	-.01	**.09**	**.16**
3	SJT-X Composite	.	.	.	.	-.02	.00	-.03	-.02	.03	.01	.05	**.12**	.00	.06	.02	.01	-.01	**.09**	-.06	-.01	-.04	.00	.02	.01	.07	.04
4	Interview Composite	.01	**.16**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5	Simulated PPW Composite	.03	**.07**	.	**.18**	.	**.23**	**.76**	**.56**	**.32**	**.19**	**.07**	**.19**	.01	**.11**	**.16**	-.01	.06	**.09**	-.03	-.05	**.13**	-.02	**.10**	**.10**	**.18**	.04
6	SimPPW Awards	-.04	**.06**	.	**.12**	**.54**	.	**.09**	**.20**	.06	.03	.06	**.32**	-.01	.08	.05	.00	**-.08**	**.08**	-.04	-.03	.01	.01	.02	-.02	.02	.03
7	SimPPW Military Education	**.07**	**.06**	.	**.11**	**.80**	**.26**	.	**.14**	.07	**.10**	.04	**.10**	-.01	.06	**.07**	-.06	-.03	.07	.00	-.03	**.12**	-.02	**.11**	**.11**	**.15**	.03
8	SimPPW Civilian Education	.00	**.09**	.	**.13**	**.50**	**.17**	**.18**	.	-.04	**.20**	-.03	.07	.05	.05	.04	**.07**	.02	-.03	.00	.00	.03	-.01	-.04	**.09**	.01	-.02
9	SimPPW Military Training	.02	-.06	.	**.11**	**.30**	**.09**	.03	-.05	.	.05	**.19**	**.27**	**-.08**	**.09**	**.21**	.02	**.19**	**.13**	-.03	.00	**.17**	**-.11**	.05	.07	**.21**	.06
10	ExAct Computer Experience	**.08**	**.14**	.	**.14**	**.22**	**.10**	**.21**	**.17**	**-.09**	.	**.08**	**.24**	**.07**	**.16**	**.17**	**.13**	.03	**.16**	-.04	**-.09**	**.16**	.01	**.23**	**.18**	**.19**	**.08**
11	ExAct Supervisory Experience	**-.13**	**-.07**	.	**.17**	**.18**	**.19**	**.06**	**.06**	**.14**	**.06**	.	**.41**	-.04	-.01	**.18**	-.05	.01	**.16**	**.08**	**.07**	.06	**-.11**	-.04	**.08**	**.21**	**-.07**
12	ExAct General Experience	-.01	.04	.	**.25**	**.36**	**.44**	**.18**	**.11**	**.29**	**.19**	**.48**	.	**-.12**	**.11**	**.21**	.04	.02	**.27**	**.09**	.03	**.24**	**-.18**	**.11**	**.17**	**.35**	.00
13	AIM Dependability	.04	**.34**	.	**.09**	**.13**	.02	**.11**	**.12**	-.03	**.11**	**-.07**	-.01	.	**.31**	**.26**	**.46**	**.20**	**.11**	**-.36**	**-.35**	-.05	**.43**	.01	-.02	.03	**.27**
14	AIM Adjustment	**.07**	**.23**	.	.06	.05	.00	.01	**.06**	**.09**	**.10**	.03	**.12**	**.30**	.	**.24**	**.48**	**.28**	**.27**	**-.33**	**-.34**	**.18**	**.17**	**.27**	**.19**	**.23**	**.42**
15	AIM Work Orientation	**.07**	**.30**	.	**.15**	**.12**	.04	.06	**.08**	**.19**	**.12**	**.18**	**.19**	**.33**	**.31**	.	**.20**	**.32**	**.59**	**-.22**	**-.31**	**.27**	**.16**	**.20**	**.16**	**.43**	**.23**
16	AIM Agreeableness	.03	**.28**	.	**.07**	.05	.02	.04	**.09**	-.01	**.08**	**-.06**	-.01	**.52**	**.43**	**.30**	.	**.18**	.06	**-.27**	**-.28**	**.09**	**.25**	.07	**.13**	.04	**.39**
17	AIM Physical Conditioning	.01	**.12**	.	**.08**	.04	-.05	-.03	.05	**.18**	.04	.03	.02	**.22**	**.24**	**.35**	**.27**	.	.01	**-.14**	**-.16**	**.12**	.05	**.14**	**.11**	**.15**	**.12**
18	AIM Leadership	**.09**	**.28**	.	**.18**	**.14**	**.09**	**.08**	.04	**.16**	**.14**	**.26**	**.31**	**.19**	**.34**	**.58**	**.12**	.03	.	**-.18**	**-.27**	**.33**	.06	**.27**	**.23**	**.58**	**.18**
19	BIQ Hostility to Authority	**-.07**	**-.30**	.	**-.08**	**-.10**	-.03	**-.10**	-.05	.00	-.02	**.11**	.04	**-.43**	**-.24**	**-.24**	**-.33**	**-.16**	**-.13**	.	**.57**	**.07**	**-.55**	**-.25**	**.13**	.02	**-.54**
20	BIQ Manipulativeness	**-.12**	**-.32**	.	**-.10**	**-.10**	.00	**-.09**	**-.10**	-.03	**-.06**	**.07**	-.01	**-.41**	**-.29**	**-.39**	**-.36**	**-.22**	**-.30**	**.59**	.	**-.12**	**-.61**	**-.34**	.00	**-.21**	**-.55**
21	BIQ Social Perceptiveness	**.09**	**.14**	.	**.16**	.04	-.01	-.01	.05	**.12**	**.18**	**.22**	**.26**	**.07**	**.20**	**.36**	**.12**	**.10**	**.41**	**.09**	**-.15**	.	**-.13**	**.24**	**.44**	**.61**	**.15**
22	BIQ Social Maturity	.03	**.26**	.	**.07**	**.11**	**.06**	**.10**	**.11**	-.05	**.06**	**-.15**	-.05	**.51**	**.15**	**.22**	**.30**	**.15**	.06	**-.62**	**-.59**	**-.08**	.	**.18**	**-.13**	**-.09**	**.43**
23	BIQ Tolerance for Ambiguity	**.18**	**.20**	.	.05	**.08**	.02	.04	**.06**	**.07**	**.13**	**.07**	**.17**	**.11**	**.33**	**.31**	**.17**	.04	**.36**	**-.35**	**-.39**	**.24**	**.22**	.	**.18**	**.25**	**.27**
24	BIQ Openness	**.11**	.05	.	**.15**	-.02	-.01	-.04	-.01	.05	**.18**	**.16**	**.22**	-.03	**.13**	**.21**	**.06**	**.07**	**.19**	**.13**	.01	**.48**	**-.11**	**.27**	.	**.44**	.03
25	BIQ Leadership	.05	**.12**	.	**.22**	**.15**	**.08**	**.07**	.05	**.19**	**.22**	**.36**	**.39**	**.07**	**.25**	**.46**	.02	**.07**	**.62**	.04	**-.18**	**.69**	**-.06**	**.31**	**.45**	.	**.12**
26	BIQ Interpersonal Skill	**.15**	**.30**	.	**.16**	**.08**	.03	**.07**	**.06**	**.06**	**.09**	**-.06**	.02	**.40**	**.34**	**.33**	**.43**	**.13**	**.25**	**-.53**	**-.55**	**.20**	**.44**	**.40**	**.12**	**.19**	.

Note. E5 Soldier correlations are presented below the diagonal (n_{E5} = 622-885). E6 Soldier correlations are presented above the diagonal (n_{E6} = 442-556). Statistically significant correlations are bolded, $p < .05$ (one-tailed).

instrument with which they are associated. Scores on instruments designed to directly address cognitive ability and skills related to judgment are shown first (i.e., ASVAB, SJT, SJT-X, and semi-structured interview), followed by measures emphasizing experience (i.e., SimPPW and its constituent scores and ExAct) and measures designed to assess temperament constructs (i.e., AIM and BIQ).

Cognitive Ability and Judgment

ASVAB. The ASVAB General Technical (GT) composite score is currently used for various post-enlistment decisions (e.g., eligibility for reenlistment) and can be considered a good measure of general cognitive aptitude. It is therefore noteworthy that its largest correlation was with the SJT composite score for E6 Soldiers ($r = .25$). This correlation for E5 Soldiers was noticeably smaller ($r = .14$). With the exception of BIQ Tolerance for Ambiguity scores ($r_{E5} = .18$, $r_{E6} = .17$), the correlations between ASVAB GT and other predictor scores were low.

SJT. As described in Chapter 6, some of the items in the E6 Soldier version of the SJT are different from those in the E4/E5 Soldier version. Although the trend was most pronounced in the E4 and E5 samples, the SJT score was related to almost all of the AIM and BIQ scales in all three samples. In the lower pay grades, the correlations with the temperament scales tended to be in the mid-.20s to the mid-.30s, whereas the highest correlation in the E6 sample was .22. When we contrast this with the findings related to ASVAB GT, it appears that judgment as measured by the SJT might be influenced more by temperament than cognitive aptitude for E4 and E5 Soldiers and relatively equally by cognitive aptitude and temperament for E6 Soldiers. These correlations with the AIM and BIQ scales imply that some aspects of personality influence Soldiers' evaluations of the best and worst ways to behave in different situations. Further research might aid the construction of an SJT that has even higher correlations with personality constructs. On the other hand, it may be that Soldiers in higher pay grades have had more training in how to handle various supervisory problems, so they rely less heavily on their own personality-driven instincts to respond to the SJT questions than their relatively less trained counterparts.

SJT-X. We administered the SJT-X to E6 Soldiers only. Its largest correlation was with the SJT ($r = .19$), and its correlations with other scores were relatively small. This makes sense, given that the SJT-X was designed to measure judgment related to a relatively narrowly defined KSA (*Knowledge of the Inter-Relatedness of Units*).

Semi-structured interview. We administered the interview to E4 and E5 Soldiers only. Generally, the results show relatively small but significant correlations between the interview composite score and other predictor scores; however, there are a few notable exceptions. The correlation between the interview and ASVAB GT was near zero for E4 and E5 Soldiers (i.e., $r = .06$ and $r = .01$, respectively). A possible explanation is that ASVAB GT measures cognitive aptitude, or a "can-do" element of the predictor space, and the interview focuses on "will-do" or "have-done" parts of the predictor space that are affected by constructs related to motivation and

temperament.[35] Consistent with this interpretation, for E5 Soldiers the interview score correlated most highly with the experience and temperament measures (e.g., ExAct General Experience, r = .25; AIM Leadership, r = .18; BIQ Leadership, r = .22). The pattern is somewhat different for E4 Soldiers, where the correlations between the interview and ExAct scores were slightly lower (e.g., ExAct General Experience, r = .19), but the correlations with the AIM scale scores were uniformly higher (e.g., AIM Leadership, r = .32) and correlations with BIQ scale scores were either comparable or higher. Similar to the results observed for the SJT, this pattern of correlations suggests that, for E5 Soldiers, variation in levels of experience (as measured by the ExAct) had a greater influence on responses to interview questions. E4 Soldiers, in contrast, may have relied more heavily on their own personality-driven instincts to respond to interview questions than did their relatively more experienced counterparts.

Promotion Point Worksheet (PPW)

When examining the correlations between the SimPPW composite score and other predictor scores, it is important to note that it is a simulation of the operational PPW. This means that the score includes operational caps that restrict the ranges and variances of its constituent scales (especially for E6 Soldiers).[36] Other scoring strategies could be considered that would result in different, possibly better, evidence for construct and criterion-related validity. For the purposes of assessing *construct* validity, the caps were not imposed on the four basic scores (i.e., SimPPW Awards, Military Education, Civilian Education, and Military Training) under the assumption that the unrestricted scores would be better measures of the underlying constructs.

Ignoring correlations with the overall SimPPW Composite, the SimPPW Awards score correlated most highly with ExAct General Experience for both E5 (r = .44) and E6 (r = .32) Soldiers. It was also correlated with ExAct Supervisory Experience for E5 Soldiers (r = .19) and even more so for E4 Soldiers (r = .31).

Correlations between the SimPPW Military Education score and scores from other instruments were generally small, though there were some relations with experience. In the E5 sample, two ExAct scores correlated with Military Education (Computer Experience, r = .21; General Experience, r = .18). This pattern was even more pronounced in the E4 sample, where Military Education correlated with all three ExAct scores (Computer Experience, r = .13; General Experience, r = .23; Supervisory Experience, r = .20). In contrast, the correlations with the experience scores for E6 Soldiers were all relatively small (r = .10 or lower).

---

[35] The job performance literature (e.g., J. Campbell & Knapp, 2001; J. Campbell, McCloy, Oppler, & Sager, 1993; Sackett, Zedeck, & Fogli, 1988) distinguishes between maximal performance (i.e., how well one can do the job when trying one's best) and typical performance (i.e., how well one will do the job—that is, how well one performs the job day in and day out). Research has demonstrated that measures of cognitive ability, perceptual speed/accuracy, and psychomotor ability show stronger relations with maximal performance measures (e.g., hands-on tests), whereas non-cognitive measures (e.g., personality/temperament constructs) show stronger relations with typical performance measures (e.g., supervisor ratings; cf. J. Campbell & Knapp, 2001; McCloy, J. Campbell, & Cudeck, 1994).

[36] See Chapter 4 for a description of the operational PPW caps.

Correlations between the SimPPW Civilian Education scores and other non-PPW scores were generally small. The only exception is ExAct Computer Experience scores for E5 and E6 Soldiers (r_{E5} = .17, r_{E6} = .20), though not for E4 Soldiers (r = .07).

The situation is different for SimPPW Military Training, where relations with scores from several other measures emerged. Chapter 4 explains that the Military Training score is a combination of the Soldier's score on a physical fitness test and a weapons qualification test. Four non-PPW predictor scores correlated relatively highly with SimPPW Military training scores for both E5 and E6 Soldiers: (a) ExAct General Experience (r_{E5} = .29, r_{E6} = .27), (b) AIM Work Orientation (r_{E5} = .19, r_{E6} = .21), (c) AIM Physical Conditioning (r_{E5} = .18, r_{E6} = .19), and (d) BIQ Leadership (r_{E5} = .19, r_{E6} = .21). The E4 sample yielded reasonably comparable results, although there was a moderately high correlation (r = .32) with ExAct Supervisory Experience. Interestingly, BIQ Social Maturity correlated -.17 with Military Training in the E4 sample and -.11 in the E6 sample. This finding suggests that Soldiers with greater social maturity tend to exhibit lower physical fitness and marksmanship skills.

Although it is not of great interest with respect to the construct validity of the PPW, the overall composite SimPPW score is included in Table 9.1 because it is the score used in subsequent criterion-related validity analyses. An examination of the correlations suggests SimPPW scores reflect somewhat different constructs across pay grades. This may be because, although the scores were computed in exactly the same way, the caps on the scales differentially impacted Soldiers in different pay grades. Thus, for example, SimPPW Awards correlated .54 with the SimPPW Composite in the E5 sample, but only .23 in the E6 sample.

Experience and Activities Record (ExAct)

As previously stated and consistent with expectations, the ExAct Computer Experience score correlated with SimPPW Civilian Education in the E5 and E6 samples, though not in the E4 sample (r_{E4} = .07, r_{E5} = .17, r_{E6} = .20). It was correlated with SimPPW Military Education for E5 Soldiers, but less so for the other two samples (r_{E4} = .13, r_{E5} = .21, r_{E6} = .10). In all three samples, computer experience also correlated with several BIQ scale scores, including Social Perceptiveness, Openness, Leadership, and Tolerance for Ambiguity. These correlations ranged from a low of .13 in the E5 sample to a high of .23 in the E6 sample (both for the correlation between ExAct Computer Experience and BIQ Tolerance for Ambiguity).

ExAct Supervisory Experience correlated relatively strongly with Leadership scores on the AIM (r_{E5} = .26, r_{E6} = .16) and BIQ (r_{E5} = .36, r_{E6} = .21). It also correlated strongly with BIQ Social Perceptiveness in the E5 sample (r = .22). Supervisory experience also correlated with AIM Work Orientation (r_{E5} = .18, r_{E6} = .18) and several other scales pertaining to initiative. Specifically, SimPPW Military Training correlated with supervisory experience in all three samples, and SimPPW Awards correlated with supervisory experience in the E4 and E5 samples. SimPPW Civilian Education also correlated with supervisory experience in the E4 sample. Interestingly, ExAct Supervisory Experience correlated negatively with ASVAB GT (r_{E5} = -.13, r_{E6} = -.14).

The ExAct General Experience score correlated highly with SimPPW Awards (r_{E5} = .44, r_{E6} = .32) and SimPPW Military Training (r_{E5} = .29, r_{E6} = .27). Appreciable correlations with other scores related to initiative are evident (see SimPPW Military Education and AIM Work Orientation). General experience related strongly to leadership as measured by AIM Leadership

(r_{E5} = .31, r_{E6} = .27) and BIQ Leadership (r_{E5} = .39, r_{E6} = .35). BIQ Social Perceptiveness (r_{E5} = .26, r_{E6} = .24) and BIQ Openness also correlated moderately (r_{E5} = .22, r_{E6} = .17). In another reassuring finding, E4 and E5 Soldier interview scores correlated with ExAct General Experience scores (r_{E4} = .19, r_{E5} = .25). While no particular relation between general experience and social maturity was hypothesized, the negative correlation between ExAct General Experience and BIQ Social Maturity scores for E6 Soldiers (r = -.18) is a bit surprising.

Temperament Measures

The AIM and BIQ instruments measure temperament constructs relevant to job performance. The scores on the two instruments show sensible relations with each other (e.g., the two leadership scales were correlated .58-.62 across the three pay grades). It is also conceptually consistent that ASVAB GT showed relatively small correlations with AIM and BIQ scale scores. For E5 and E6 Soldiers, the highest correlation was between ASVAB GT and BIQ Tolerance for Ambiguity scores (r_{E5} = .18, r_{E6} = .17). As mentioned previously, the correlations suggested substantial relations between SJT and the AIM and BIQ scores that were somewhat stronger for E5 than E6 Soldiers (e.g., SJT with AIM Dependability, r_{E5} = .34, r_{E6} = .20). Generally, AIM and BIQ scale scores had relatively low correlations with the PPW scale scores. The exception was SimPPW Military Training with AIM Work Orientation (r_{E5} = .19, r_{E6} = .21), AIM Physical Conditioning (r_{E5} = .18, r_{E6} = .19), and BIQ Leadership (r_{E5} = .19, r_{E6} = .21). All ExAct Supervisory and General Experience scale scores had relatively strong correlations with AIM Work Orientation, AIM Leadership, and BIQ Leadership for E5 and E6 Soldiers (e.g., AIM Leadership with ExAct General Experience, r_{E5} = .31, r_{E6} = .27). Other notable relations with the ExAct scores include all three ExAct scores with BIQ Social Perceptiveness for E5 Soldiers (i.e., ExAct Computer Experience, r = .18; ExAct Supervisory Experience, r = .22; ExAct General Experience, r = .26). Finally, the negatively stated BIQ scale scores (i.e., BIQ Hostility to Authority and BIQ Manipulativeness) showed expected and logical negative correlations with a number of other scores (e.g., BIQ Hostility to Authority with SJT, r_{E5} = -.30; BIQ Manipulativeness with AIM Dependability, r_{E5} = -.43; BIQ Manipulativeness with AIM Dependability, r_{E6} = -.36).

Summary

Taken together, the correlations among the NCO21 predictor scores show patterns that are consistent with expectations and provide evidence of their construct validity. The correlations also provide some interesting insight into the individual difference constructs that the predictors assess. For example, the variables related to SJT performance appear to differ across pay grade. The low correlations between general cognitive ability, as assessed by the ASVAB GT, and other variables are reassuring in the sense that the other measures are tapping something considerably distinct from *g*.

The finding that SimPPW Military Training correlates with other variables suggests a common "motivation" factor. Indeed, we conducted some factor analyses of the predictor scores in an attempt to identify underlying common factors but could not arrive at an interpretable solution that was not dominated by method factors. We did not aggressively pursue this course (as we did with the criterion scores) because, given our operational testing goals, we are more interested in the meaning of the actual predictor scores than the underlying factors they might represent.

The most salient unexpected result was the BIQ Social Maturity scale's negative correlations with SimPPW Military Training and ExAct General Experience. These correlations are most likely due to a confound with gender. Subgroup difference tables in Chapters 4, 5, and 8 show that BIQ Social Maturity scores were significantly higher for females compared to males in all pay grades ($p < .001$) and SimPPW Military Training and ExAct General Experience scores were significantly higher for males compared to females in all pay grades ($p < .001$).

Relations between Predictor Scores and Observed Performance Scale-Level Ratings

Additional evidence for the construct validity of the predictors can be obtained by examining the pattern of correlations between the predictor scores and the individual performance scale ratings (see Tables 9.2 and 9.3). Some of the predictors were designed to assess KSAs that serve as determinants of one or more performance dimensions. For example, the AIM Leadership score was developed to assess personality trait characteristics that should predict performance on the Leadership performance dimensions. Therefore, a high correlation between the AIM Leadership score and the Leadership performance ratings would be considered evidence supporting the construct validity of this potential predictor. Similarly, the AIM Dependability score should be related to the integrity, discipline, and self-management dimension ratings. If the pattern of correlations is consistent with these theoretical expectations, it would be evidence in support of the construct validity of the AIM Dependability measure. This presumes a measure of construct validity for the performance rating scales as well. If the pattern is not as expected, it is not necessarily because the AIM does not possess construct validity. Questions about the construct validity of the ratings might be more reasonable.

The SJT correlated significantly with all of the observed performance scales for E5 Soldiers (see Table 9.2). Thus, the SJT predicts the full spectrum of performance as assessed by these supervisor ratings. Its highest correlations were with the leadership and peer support performance scales (e.g., Relating to and Supporting Peers) rather than problem-solving or information-related performance scales (e.g., Problem-Solving/Decision-Making Skill). These results were consistent with the SJT's correlations with the other predictors. However, the pattern of correlations did not perfectly match up with the KSAs the SJT was designed to measure. Among E6 Soldiers, only 11 of 18 correlations with the observed performance scales were significant (see Table 9.3). The significant correlations were mostly with the cognitive-task-related scales. This finding was consistent with the higher correlation between the SJT and the ASVAB GT for E6 Soldiers ($r = .25$) vs. E5 Soldiers ($r = .14$) shown in Table 9.1. When evaluating these results in terms of the SJT's construct validity, remember that (a) the final selection of items was based on their relationship with the observed performance composite and (b) half of the items in the E6 Soldier version of the SJT are different from those in the E4/E5 Soldier version.

The SJT-X and SJT had similar patterns of correlations with the observed performance scales. The SJT-X, however, had substantially higher correlations than the SJT with Common Task Knowledge and Skill, Adaptability, and Leadership Skills. Correlations were also computed between the SJT-X and other measures (i.e., individual items from the instruments) that were more closely related to the constructs that the SJT-X was designed to measure. These relationships are discussed in Chapter 6.

Table 9.2. Raw Correlations between Predictor Scores and Observed Performance Scale-Level Ratings for E5 Soldiers

Predictors/Criterion Scale	MOS/Occupation-Specific Knowledge and Skill	Common Task Knowledge and Skill	Computer Skills	Writing Skill	Oral Communication Skill	Level Of Effort/Initiative on the Job	Adaptability	Self-Management and Self-Directed Learning Skill	Demonstrated Integrity, Discipline, etc.	Acting as a Role Model	Relating to and Supporting Peers	Cultural Tolerance	Selfless Service Orientation	Leadership Skills	Concern for Soldier Quality of Life	Training Others	Problem-Solving/Decision Making Skill	Information Management
ASVAB GT	.04	**.08**	**.12**	**.14**	.05	.01	.02	.06	.03	.04	.03	-.01	**.08**	.00	.06	.02	.06	**.10**
SJT Composite	**.17**	**.18**	**.13**	**.15**	**.12**	**.16**	**.17**	**.15**	**.20**	**.15**	**.21**	**.08**	**.19**	**.18**	**.19**	**.17**	**.13**	**.13**
SJT-X Composite	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Interview Composite	**.13**	**.09**	.06	**.10**	**.15**	**.15**	**.14**	**.17**	**.10**	**.13**	**.10**	**.10**	**.11**	**.17**	.06	**.12**	**.16**	.03
SimPPW Awards	**.18**	**.12**	**.08**	.04	.03	.06	**.08**	.00	-.01	-.02	.05	.02	.04	**.09**	.05	**.13**	**.12**	.06
SimPPW Military Education	**.19**	**.10**	**.14**	**.15**	**.12**	**.10**	**.09**	**.14**	**.15**	**.11**	**.11**	.06	**.10**	**.09**	**.10**	**.14**	**.13**	**.12**
SimPPW Civilian Education	.04	.01	**.13**	**.08**	-.01	.03	.02	**.11**	-.01	.04	.01	.00	.03	.05	.06	.04	.01	**.08**
SimPPW Military Training	**.15**	**.18**	.03	**.12**	**.16**	**.15**	**.14**	**.14**	**.09**	**.21**	**.09**	.06	**.10**	**.22**	**.12**	**.17**	**.12**	**.09**
ExAct Computer Experience	**.10**	.06	**.36**	**.20**	.07	.00	.00	.05	.01	-.03	.01	**.07**	.02	.03	.07	.00	.05	**.10**
ExAct Supervisory Experience	**.11**	**.13**	-.01	.05	**.08**	.06	**.08**	.05	.03	.00	.03	-.03	.05	**.14**	**.09**	**.08**	**.10**	.00
ExAct General Experience	**.19**	**.23**	.05	.07	**.12**	**.07**	**.11**	.02	.05	.01	.06	.04	**.10**	**.15**	**.08**	**.17**	**.15**	.03
AIM Dependability	.04	.05	**.08**	**.07**	**.07**	.05	.04	**.12**	**.17**	**.09**	**.10**	**.10**	.05	.06	**.12**	.06	.05	.02
AIM Adjustment	.03	.02	.03	.03	.05	.04	**.10**	.00	.01	.01	**.08**	**.10**	.04	.06	.06	.01	**.08**	.02
AIM Work Orientation	**.23**	**.19**	**.09**	**.16**	**.17**	**.26**	**.25**	**.20**	**.19**	**.22**	**.20**	**.12**	**.21**	**.29**	**.21**	**.20**	**.19**	**.13**
AIM Agreeableness	-.02	-.06	-.01	-.01	-.02	.00	.05	.03	**.07**	.04	**.08**	**.08**	-.01	-.01	.04	-.02	.00	-.02
AIM Physical Conditioning	.06	**.09**	.04	**.09**	.07	.06	.05	**.08**	**.07**	**.15**	**.07**	**.08**	**.07**	**.12**	**.07**	.06	.07	.03
AIM Leadership	**.17**	**.16**	**.09**	**.20**	**.23**	**.15**	**.18**	**.15**	**.12**	**.09**	**.15**	**.07**	**.12**	**.25**	**.18**	**.17**	**.18**	**.11**
BIQ Hostility to Authority	-.03	-.03	-.06	-.04	-.04	-.04	-.03	**-.07**	-.02	-.07	-.02	-.04	-.05	-.05	-.04	-.03	-.06	-.05
BIQ Manipulativeness	-.05	-.01	**-.07**	**-.10**	-.05	-.06	**-.07**	**-.10**	**-.07**	**-.08**	-.04	-.02	-.07	-.04	-.04	-.05	-.03	-.01
BIQ Social Perceptiveness	**.10**	**.14**	**.11**	**.15**	**.20**	.07	**.12**	**.08**	**.10**	**.08**	**.11**	.05	.05	**.13**	**.12**	**.12**	**.12**	.04
BIQ Social Maturity	.02	-.02	**.10**	.07	.01	.06	.01	.06	**.10**	**.07**	.02	.05	.07	.01	**.08**	-.02	.01	.01
BIQ Tolerance for Ambiguity	**.15**	**.10**	**.12**	**.13**	**.11**	**.12**	**.08**	**.09**	.06	.06	**.10**	.06	**.08**	**.12**	**.09**	**.11**	**.12**	**.08**
BIQ Openness	-.01	.04	.03	.05	.04	.05	.05	.01	.07	.03	.04	.05	.00	.04	.02	.01	.06	-.01
BIQ Leadership	**.19**	**.19**	**.12**	**.19**	**.24**	**.17**	**.20**	**.17**	**.13**	**.12**	**.16**	**.09**	**.13**	**.25**	**.18**	**.22**	**.21**	**.08**
BIQ Interpersonal Skill	.06	.06	**.11**	**.13**	**.11**	.06	**.09**	**.08**	**.07**	**.07**	**.12**	**.07**	.04	.04	.05	.06	**.09**	**.08**

Note. n = 471-608. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

Table 9.3. Raw Correlations between Predictor Scores and Observed Performance Scale-Level Ratings for E6 Soldiers

Predictors/Criterion Scale	MOS/Occupation-Specific Knowledge and Skill	Common Task Knowledge and Skill	Computer Skills	Writing Skill	Oral Communication Skill	Level Of Effort/Initiative on the Job	Adaptability	Self-Management and Self-Directed Learning Skill	Demonstrated Integrity, Discipline, etc.	Acting as a Role Model	Relating to and Supporting Peers	Cultural Tolerance	Selfless Service Orientation	Leadership Skills	Concern for Soldier Quality of Life	Training Others	Problem-Solving/Decision Making Skill	Information Management
ASVAB GT	.08	.07	**.17**	**.20**	**.13**	.05	.01	.04	.06	-.01	.00	.06	**.10**	.05	.05	.06	.03	**.16**
SJT Composite	**.11**	**.09**	**.22**	**.16**	**.15**	**.09**	.04	.05	**.10**	.06	.08	.07	**.17**	.05	.08	**.11**	**.15**	**.19**
SJT-X Composite	**.09**	**.16**	.02	**.10**	**.18**	.05	**.13**	.06	-.01	.05	.07	**.12**	.08	**.16**	.02	**.14**	**.16**	**.10**
Interview Composite	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SimPPW Awards	**.13**	**.16**	.05	.07	**.14**	**.12**	**.16**	**.11**	.06	.01	**.11**	**.10**	**.15**	**.15**	**.11**	**.19**	**.17**	**.09**
SimPPW Military Education	.09	.05	.05	-.01	.00	.00	.06	-.01	.01	.05	-.06	-.06	-.05	.03	**-.09**	.05	.05	.04
SimPPW Civilian Education	**.09**	.02	**.13**	**.12**	.03	.03	.06	**.09**	-.01	.04	.02	.06	.01	-.03	.03	.07	.07	**.11**
SimPPW Military Training	**.09**	**.13**	.01	-.03	.01	.01	.07	.02	.01	**.20**	.00	.00	.00	**.11**	-.05	.08	.04	-.02
ExAct Computer Experience	.01	.07	**.19**	**.09**	.03	-.03	.07	**.10**	.01	.03	.03	.02	.01	.00	.06	-.02	.05	**.12**
ExAct Supervisory Experience	-.02	.06	-.02	**-.12**	-.04	-.07	.03	-.02	-.09	.07	-.01	-.02	-.03	.04	.02	.03	-.04	-.05
ExAct General Experience	.07	**.15**	.05	-.02	**.09**	-.08	.06	.01	-.08	.01	.06	**.12**	-.02	.08	.08	.07	**.12**	.07
AIM Dependability	-.06	-.04	.02	-.01	-.01	-.01	-.04	.04	.06	.03	.03	.03	.03	-.07	.00	-.02	-.07	-.02
AIM Adjustment	**.11**	**.10**	.07	.07	.00	-.05	.06	.04	.02	**.13**	-.01	.01	.04	.06	.03	.01	.04	**.09**
AIM Work Orientation	**.11**	**.15**	-.02	.03	-.01	**.11**	.04	**.14**	.03	**.14**	.00	.00	.06	**.15**	-.01	**.11**	.04	.02
AIM Agreeableness	-.03	.02	.05	.00	-.04	**-.10**	.00	-.02	.02	.03	.04	.06	.04	-.04	.03	-.04	-.06	.00
AIM Physical Conditioning	.06	.08	-.03	-.01	-.06	-.03	.01	.03	-.01	**.15**	.01	.03	-.03	.07	-.01	.02	-.01	.01
AIM Leadership	**.10**	**.11**	-.01	.05	.03	.07	.06	.08	.03	.05	-.04	-.04	.07	**.11**	-.07	.04	.02	.05
BIQ Hostility to Authority	**-.10**	**-.09**	-.05	**-.12**	**-.12**	-.08	-.06	-.07	**-.16**	**-.14**	**-.10**	**-.10**	**-.11**	**-.14**	-.02	-.03	.02	-.05
BIQ Manipulativeness	-.02	-.06	-.08	**-.13**	**-.14**	-.08	-.06	**-.11**	**-.13**	-.06	.03	-.07	-.04	**-.11**	.01	-.07	-.03	**-.09**
BIQ Social Perceptiveness	.05	.00	-.06	-.08	-.03	-.02	.03	-.05	-.07	.00	-.04	-.05	-.01	.05	-.04	.08	.03	.02
BIQ Social Maturity	-.03	-.02	**.14**	**.11**	.08	.06	.01	**.10**	**.09**	.00	.00	.06	.03	.01	.04	.02	-.03	.01
BIQ Tolerance for Ambiguity	**.10**	.06	-.01	**.12**	.02	.02	.00	.04	.03	-.02	-.01	-.02	.05	.07	-.06	.05	-.02	**.10**
BIQ Openness	-.01	-.09	-.03	**-.09**	**-.12**	-.04	-.07	**-.09**	.01	-.05	-.01	.00	-.03	-.07	-.06	.03	-.04	-.01
BIQ Leadership	**.15**	.04	-.04	-.04	.02	.04	.08	.01	-.07	.02	-.04	-.04	.02	**.09**	.03	**.09**	.01	.05
BIQ Interpersonal Skill	**.09**	.07	**.11**	**.10**	**.10**	**.09**	.04	**.10**	**.11**	**.14**	.07	**.14**	**.13**	**.11**	.05	**.09**	.07	.05

Note. n = 341-393. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

The semi-structured interview yields the following nine scores (each of which represents a KSA or set of KSAs): Adaptability, Level of Effort and Initiative on the Job, Level of Integrity and Discipline on the Job, Relating to and Supporting Peers, Leadership Skills/Potential, Oral Communication Skill, Self-Management and Self-Directed Learning Skill, MOS/Occupation-Specific Knowledge and Skill, and Military Presence. Conceptually, each of these scores (except for Military Presence) directly relates to an observed performance scale. Although the composite interview score did correlate highest with the eight observed performance scales it was intended to predict (see Table 9.2), the expected pattern based on the individual interview rating scales did not emerge. The composite score also correlated significantly with Problem-Solving/Decision Making Skill, although the interview does not produce a score related to this performance scale.

The ExAct Computer Experience score's highest correlation, by far, was with the Computer Skills performance scale ($r_{E5} = .36$, $r_{E6} = .19$). For E5 Soldiers, the ExAct Supervisory Experience score correlated highest with performance scales related to supervision. This score had no significant correlations for E6 Soldiers (except for a negative correlation with Writing Skill). The correlation patterns were similar for the ExAct General Experience score. This score, however, predicted the two knowledge and skill performance scales much better than did ExAct Supervisory Experience.

Two AIM scores—Work Orientation and Leadership—had the highest correlations with the performance scales (and the most significant correlations). For the E5 Soldiers, each AIM score correlated highest with the performance scale that is conceptually most relevant. For the AIM Physical Conditioning score, no observed performance scale was directly related. It does make sense, however, that a good Army role model and leader would have good physical conditioning, which is consistent with the correlations. For E6 Soldiers, the correlations were much lower, and the patterns of the AIM scores' correlations with the performance scales were somewhat different.

Among E5 Soldiers, the BIQ Leadership score correlated highest with the Leadership Skill performance scale. Most of the other BIQ scores had significant (but not high) correlations with several observed performance scales. These correlations were generally consistent with the expected relationships between the BIQ scores and the observed performance scales. However, the BIQ Openness score had no significant correlations with observed performance scales, and the BIQ Hostility to Authority score had only one. The results differed somewhat for E6 Soldiers. In particular, BIQ Social Perceptiveness had no significant correlations with the observed performance scales, whereas BIQ Hostility to Authority had 10.

One other finding is worth noting: Although oral communication skill was directly measured only by the interview, it correlated moderately with both the BIQ and AIM Leadership scores (see Table 9.2). Thus, it appears that E5 Soldiers who obtain high Leadership scores also tend to have good oral communication skills.

In summary, several points can be made regarding the results for E5 Soldiers. The interview, which was designed to measure skills directly related to observed performance scales, exhibited significant estimates of criterion-related validity for the composite score but no discriminant validity for the individual scales. The AIM Leadership and BIQ Leadership scores showed clear evidence of construct validity. The ExAct's correlations with the observed performance scales exhibited some evidence of construct validity (particularly for the ExAct Computer Experience score). The SJT's correlations with all of the observed performance scales

were consistent with its heterogeneous nature and somewhat supportive of its construct validity, given the KSAs the SJT was designed to measure. For the remaining scores, it is more difficult to evaluate their construct validity based on their correlations with the observed performance scales. These scores tended to correlate with several performance scales; however, no correlations were directly inconsistent with the expected relations.

The correlations were generally lower for E6 Soldiers. Not only were the values smaller, but the patterns of correlations differed somewhat, as well. As described later, with the exception of ASVAB GT, the correlations between the predictors and the composite criteria measures were also lower for E6 Soldiers compared to E5 Soldiers. However, it is important to note that the construct validity of some of these predictors (e.g., ASVAB GT) has been supported in previous research (e.g., J. Campbell & Knapp, 2001). Therefore, different relations between predictors and criteria across pay grades in this effort may reflect substantive pay grade differences in the determinants of performance (at least as assessed by supervisor raters).

Relations between Predictor Scores and Criterion Factor Scores

The relations between the predictor measures and job performance were examined in another way, as well. To simplify the interpretation of the predictor-criterion relations, the 6 observed performance factors (as described in Chapter 3) were used rather than the 18 observed performance rating scales (see Table 9.4). Construct validity of the predictor measures was easier to assess using the factor scores: Technical Performance, Leadership Structure, Effort/Integrity/Selfless Service, Leadership Consideration, Information Management, and Individual Self-Management.

Two general differences between E5 and E6 Soldiers can be seen. First, the correlations were lower for E6 Soldiers than for E5 Soldiers. This is consistent with the preceding analyses. Second, for E5 Soldiers, the AIM and BIQ had higher correlations with the performance factors than did the ASVAB GT. In contrast, for E6 Soldiers, correlations between the AIM and BIQ scales and the performance factors were generally lower than their correlations with the ASVAB GT. This implies that E5 and E6 Soldiers differ in the relative impact of general cognitive ability vs. personality as determinants of job performance at the two levels.

ASVAB GT correlated significantly with the two performance factors expected to be most directly related to general cognitive ability: Technical Performance and Information Management. Indeed, this relation was a little stronger for E6 Soldiers than for E5 Soldiers. The SJT and interview composite scores, designed to measure a variety of KSAs, correlated significantly with all six performance factors for E5 Soldiers. The SJT showed a similar pattern of correlations for E6 Soldiers. Taken together, the patterns of these correlations support the construct validity of these predictors.

Performance factor correlations with the PPW were lower for E6 Soldiers than for E5 Soldiers. Among E5 Soldiers, SimPPW Military Training and Military Education correlated significantly with all six performance factors. SimPPW Civilian Education primarily predicted Information Management. Among E6 Soldiers, SimPPW Awards correlated significantly with four performance factors. These correlations represent moderate support for the construct validity of the PPW scales scores. (See this project's recommendations report [Knapp et al., 2003] for a discussion of potential improvements to the operational PPW.)

Table 9.4 Raw Correlations between Predictor Scores and Criterion Factor Scores by Pay Grade

Predictor	E5 Soldiers						E6 Soldiers					
	TPF	LDS	EIS	LDC	INF	SLF	TPF	LDS	EIS	LDC	INF	SLF
ASVAB GT Score	.07	.04	.05	.04	**.15**	.05	**.09**	.07	.08	.04	**.21**	.01
SJT Composite	**.19**	**.19**	**.21**	**.21**	**.17**	**.17**	**.11**	**.12**	**.14**	**.10**	**.23**	.06
SJT-X Composite	-	-	-	-	-	-	**.14**	**.19**	.05	**.09**	.09	.06
Interview Composite	**.12**	**.18**	**.14**	**.11**	**.08**	**.17**	-	-	-	-	-	-
SimPPW Awards	**.17**	**.11**	.04	.05	**.08**	-.01	**.16**	**.20**	**.13**	**.14**	.08	.06
SimPPW Military Education	**.16**	**.14**	**.13**	**.11**	**.17**	**.14**	.08	.05	-.02	**-.09**	.03	.03
SimPPW Civilian Education	.02	.03	.02	.03	**.12**	**.08**	.06	.05	.01	.04	**.15**	.07
SimPPW Military Training	**.18**	**.20**	**.13**	**.11**	**.10**	**.20**	**.12**	.08	.01	-.02	-.01	**.14**
ExAct Computer Experience	**.09**	.03	.01	.06	**.28**	.01	.04	.03	-.01	.05	**.17**	.07
ExAct Supervisory Experience	**.14**	**.12**	.06	.04	.02	.03	.02	.01	-.08	-.01	-.07	.04
ExAct General Experience	**.23**	**.17**	**.08**	**.08**	.06	.02	**.12**	**.10**	-.07	**.11**	.04	.01
AIM Dependability	.05	**.07**	**.11**	**.13**	**.07**	**.12**	-.06	-.06	.03	.03	-.01	.04
AIM Adjustment	.03	**.07**	.04	**.10**	.04	.01	**.12**	.04	.00	.01	**.09**	**.11**
AIM Work Orientation	**.23**	**.27**	**.26**	**.22**	**.16**	**.23**	**.14**	**.09**	**.08**	-.01	.01	**.16**
AIM Agreeableness	-.04	.00	.02	**.08**	-.01	.03	-.01	-.04	-.02	.06	.02	.01
AIM Physical Conditioning	**.08**	**.09**	**.08**	**.09**	.07	**.13**	.08	.01	-.03	.01	-.01	**.11**
AIM Leadership	**.18**	**.25**	**.15**	**.17**	**.17**	**.13**	**.12**	.07	.07	-.06	.03	.07
BIQ Hostility to Authority	-.03	-.05	-.04	-.04	-.06	**-.08**	**-.10**	-.08	**-.13**	**-.09**	**-.09**	**-.12**
BIQ Manipulativeness	-.04	-.06	**-.08**	-.04	**-.07**	**-.10**	-.05	**-.10**	**-.10**	-.01	**-.12**	**-.10**
BIQ Social Perceptiveness	**.13**	**.17**	**.08**	**.12**	**.13**	**.09**	.03	.04	-.04	-.05	-.05	-.03
BIQ Social Maturity	.00	.01	**.09**	.06	**.08**	**.07**	-.03	.02	.07	.03	**.11**	.05
BIQ Tolerance for Ambiguity	**.14**	**.13**	**.10**	**.11**	**.13**	**.08**	.09	.03	.04	-.04	.08	.01
BIQ Openness	.02	.05	.05	.04	.03	.02	-.05	-.07	-.03	-.03	-.05	-.08
BIQ Leadership	**.21**	**.28**	**.17**	**.18**	**.17**	**.16**	**.10**	.07	.00	-.02	-.01	.02
BIQ Interpersonal Skill	**.07**	**.09**	**.07**	**.10**	**.13**	**.09**	**.09**	**.10**	**.13**	**.11**	**.11**	**.14**

Note. TPF = Technical Performance; LDS = Leadership: Structure; EIS = Effort/Integrity/Selfless Service; LDC = Leadership: Consideration; INF = Information Management; SLF = Individual Self-Management. n_{E5} = 471-608, n_{E6} = 341-391. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

The construct validity of the ExAct Computer Experience score also received support. For both pay grades, it correlated relatively highly only with the Information Management factor. The ExAct Supervisory Experience score correlated significantly with Leadership Structure (and Technical Performance), but not Leadership Consideration for E5 Soldiers; it showed no significant correlations for E6 Soldiers. ExAct General Experience correlated significantly with both Leadership factors and Technical Performance for both pay grades. Overall, the correlations offer some support for the construct validity of the ExAct scales (particularly for ExAct Computer Experience).

Most correlations involving the AIM were consistent with the conceptual relations between the AIM scores and the performance factors for E5 Soldiers. For example, Agreeableness correlated significantly with only Leadership Consideration. For E5 Soldiers at least, the pattern of correlations between the predictors and performance factors is evidence of the construct validity of the AIM.

Among E5 Soldiers, BIQ Leadership correlated highest with the Leadership Structure factor. This is evidence for the construct validity of the BIQ Leadership score. The BIQ Interpersonal Skill score correlated significantly with all six performance factors for both E5 and E6 Soldiers. The related BIQ score, Social Perceptiveness, also correlated significantly with all six performance factors, but only for E5 Soldiers. Again, the pattern of correlations provides evidence supporting the validity of BIQ scores as predictors of performance, as measured by supervisor ratings, for E5 Soldiers. However, the evidence is somewhat weaker for E6 Soldiers.

Summary: Construct Validity

In general, there was good evidence for the (a) construct validity of most of the predictor measures and (b) use of the predictor measures as predictors of job performance. The best evidence of construct validity was for the leadership predictor scores: AIM Leadership and BIQ Leadership. The evidence was mixed for other scores. For example, some of the BIQ Social Maturity score's correlations (or lack of significant correlations) were unexpected, whereas some of its other correlations were very consistent with its conceptual meaning. Our best explanation for this result is the gender confound mentioned earlier. Although the BIQ Openness score did not relate significantly to any of the performance scales, it did have significant correlations with several conceptually related predictors. The measures with mixed results may be omitting some aspects of their theoretical construct domain (where expected relationships with other measures are missing) or adding aspects of foreign constructs (where unexpected relationships with other measures exist). It is also the case, however, that the AIM and BIQ were pre-existing measures not designed to measure NCO21 KSAs, per se.

Correlations among the predictors supported the construct validity of the interview and the SJT, and the relations between these scores and the criteria showed that they were related to our measures of job performance. However, the pattern of relations between the (a) interview and SJT scores and (b) specific observed performance rating scales and factors offered more equivocal support of construct validity. The ExAct's correlations with other predictors and criteria exhibited some evidence of construct validity (particularly for the ExAct Computer Experience score).

The criterion-predictor correlations were understandably weaker than the predictor intercorrelations because the KSA constructs are not perfectly related to the performance constructs. Of course, other factors (e.g., predictors are self-report whereas the performance measures are completed by Soldiers' supervisors) could also be attenuating these correlations.

The E6 sample exhibited correlations that were (a) slightly lower for the predictor interrelations and (b) much lower for the predictor-criterion relations. However, as mentioned earlier, (a) the construct validity of some of these predictors (e.g., ASVAB GT) has been supported in previous research, and (b) different relations between predictors and criteria across pay grades in this effort may reflect pay grade differences in the determinants of performance as measured by ratings, rather than problems with the construct validity of the predictors.

Criterion-Related Validity

The primary question addressed by this project is, "Which predictors will be most valid for predicting the future performance of E5 and E6 Soldiers?" Although comparing the magnitude of zero-order validity estimates is a useful heuristic for making such a determination (as was done in preceding chapters), other indices are also useful. For example, given the existing semi-centralized NCO promotion system, one useful index would be the incremental validity of each predictor over the criterion-related validity of the current system. In this investigation, the criterion-related validity of the SimPPW Composite provided an estimate of the criterion-related validity of the current promotion process and thus provided a basis for examining the incremental validity of each of the other predictors considered separately and together. The three sections that follow summarize the results of these analyses.

Although we discuss finding related to ratings of both current and expected future performance, it is not clear how well supervisors could really distinguish between the two. Indeed rating of future performance are probably driven largely by the raters' perceptions of current performance and they probably should be. But this phenomenon makes it dangerous to draw strong conclusions about differences in how well the experimental predictors truly relate to performance in the future versus current job performance. As shown in Chapter 3, the current and future rating are fairly highly correlated (r = .81 -.82).

Zero-Order Validity Estimates

Table 9.5 presents raw and corrected validity estimates for each predictor score. Although this information is available in tables presented in each predictor-specific chapter of this report, we now present these results together to aid in cross-instrument comparisons. This discussion will focus on the corrected validity estimates.

In general, the validity estimates were higher for E5 than for E6 Soldiers. This finding is discussed at the end of this chapter. The ASVAB GT, in contrast, had much higher validity estimates for E6 Soldiers than for E5 Soldiers. This finding, which was mentioned earlier in this chapter, suggests that general cognitive ability had a greater impact on E6 level performance than on E5 level performance.

Table 9.5. Raw and Corrected Correlations between Predictor and Criterion Scores by Pay Grade

Predictors	Raw				Corrected			
	Observed Performance Composite		Expected Future Performance Composite		Observed Performance Composite		Expected Future Performance Composite	
	E5	E6	E5	E6	E5	E6	E5	E6
SimPPW Composite	**.19**	**.09**	**.11**	**.11**	.19	.13	.13	.18
ASVAB GT Score	**.08**	**.11**	.06	**.10**	.11	.19	.10	.20
SJT Composite	**.23**	**.16**	**.19**	**.16**	.39	.25	.37	.28
SJT-X Composite	.	**.14**	.	**.15**	.	.18	.	.22
Interview Composite	**.17**	.	**.15**	.	.25	.	.26	.
ExAct Computer Experience	**.09**	.07	**.08**	**.12**	.14	.10	.14	.21
ExAct Supervisory Experience	**.08**	-.02	**.11**	.03	.21	-.03	.30	.05
ExAct General Experience	**.13**	.07	**.12**	.06	.19	.10	.20	.11
AIM Dependability	**.11**	-.01	**.12**	.01	.17	-.02	.21	.02
AIM Adjustment	.06	.07	.05	**.12**	.08	.10	.08	.19
AIM Work Orientation	**.28$_a$**	**.09**	**.28$_a$**	**.11**	.40	.13	.46	.17
AIM Agreeableness	.01	-.01	-.01	.02	.02	-.01	-.02	.03
AIM Physical Conditioning	**.11**	.02	**.10**	.04	.15	.03	.16	.06
AIM Leadership	**.22$_a$**	.06	**.26$_a$**	.08	.33	.09	.43	.12
BIQ Hostility to Authority	-.06	**-.13**	**.07**	**-.10**	-.08	-.17	-.11	-.15
BIQ Manipulativeness	**-.08**	**-.10**	**-.07**	**-.10**	-.11	-.15	-.11	-.17
BIQ Social Perceptiveness	**.15$_a$**	-.01	**.16$_a$**	.03	.21	-.02	.25	.04
BIQ Social Maturity	.06	.06	.01	.07	.09	.08	.02	.11
BIQ Tolerance for Ambiguity	**.14**	.04	**.13**	.08	.18	.07	.19	.14
BIQ Openness	.05	-.06	.07	-.05	.06	-.09	.10	-.08
BIQ Leadership	**.25$_a$**	.04	**.27$_a$**	.06	.33	.05	.42	.09
BIQ Interpersonal Skill	**.11**	**.14**	**.09**	**.14**	.16	.18	.15	.21

Note. n_{E5} = 471-613; n_{E6} = 341-399. "Corrected" correlations were corrected for criterion unreliability and range restriction on the predictor. The "a" subscripts on E5 correlations indicate that corresponding E5 and E6 correlations were significantly different from each other, $p < .05$ (two-tailed). Statistically significant correlations are bolded, $p < .05$ (one-tailed).

According to the corrected correlations, the ExAct, AIM, and BIQ were generally slightly better at predicting expected future performance than observed performance. This might be because supervisor raters are basing their future predictions on temperament and experience (e.g., being able to count on a Soldier now means I can count on him in the future). The simulated PPW score, however, predicted the observed performance of E5 Soldiers better than their expected future performance.

Two predictors exhibited an interaction between pay grade and observed vs. future ratings. Among E5 Soldiers, the simulated PPW did better at predicting observed performance; for E6 Soldiers, it did better at predicting expected future performance. Similarly, ExAct Computer Experience was a better predictor of expected future performance than of observed performance for E6 Soldiers but not for E5 Soldiers.

Overall, the ExAct, BIQ, and AIM scores had lower validity estimates than the SJT[37], SJT-X, and interview. One BIQ score (Leadership) and two AIM scores (Work Orientation and Leadership), however, had the highest validity estimates (at least in the E5 sample).

Incremental Validity Estimates

Table 9.6 shows incremental validity estimates (over SimPPW) calculated on raw and corrected predictor score-criterion composite correlation matrices. This discussion will focus on the corrected validity estimates.

Table 9.6. Incremental Validity Estimates of Predictors Scores beyond the Simulated PPW Composite by Pay Grade

Predictors	Raw				Corrected			
	Observed Performance Composite		Expected Future Performance Composite		Observed Performance Composite		Expected Future Performance Composite	
	E5	E6	E5	E6	E5	E6	E5	E6
ASVAB GT Score	.01	**.06**	.01	**.05**	.04	.04	.02	.03
SJT Composite	**.09**	**.10**	**.10**	**.09**	.20	.09	.26	.09
SJT-X Composite	.	**.07**	.	**.07**	.	.06	.	.08
Interview Composite	**.05**	.	**.07**	.	.16	.	.24	.
ExAct Computer Experience	.01	.02	.01	**.05**	.02	.00	.04	.06
ExAct Supervisory Experience	.01	.01	**.03**	.00	.03	.04	.00	.00
ExAct General Experience	.01	.01	**.03**	.01	.00	.00	.01	.01
AIM Dependability	.01	.00	**.04**	.00	.06	.01	.17	.00
AIM Adjustment	.01	.02	.01	**.06**	.03	.00	.06	.08
AIM Work Orientation	**.13**	.03	**.18**	.03	.26	.00	.45	.02
AIM Agreeableness	.00	.00	.00	.00	.02	.01	.01	.02
AIM Physical Conditioning	**.02**	.00	**.03**	.01	.11	.00	.22	.00
AIM Leadership	**.08**	.02	**.15**	.02	.18	.00	.34	.01
BIQ Hostility to Authority	.01	**.06**	.02	.03	.01	.03	.10	.01
BIQ Manipulativeness	.01	**.04**	.02	.04	.03	.04	.08	.03
BIQ Social Perceptiveness	**.05**	.00	**.09**	.00	.13	.01	.18	.01
BIQ Social Maturity	.00	.02	.00	.02	.01	.00	.00	.00
BIQ Tolerance for Ambiguity	**.04**	.01	**.06**	.02	.09	.00	.08	.01
BIQ Openness	.01	.02	.02	.02	.00	.01	.02	.01
BIQ Leadership	**.11**	.00	**.18**	.01	.23	.01	.38	.00
BIQ Interpersonal Skill	**.03**	**.07**	**.03**	**.06**	.10	.03	.16	.03

Note. n_{E5} = 469-603; n_{E6} = 337-395. "Raw" coefficients were calculated on the uncorrected correlation matrix of predictors and the criterion. "Corrected" coefficients were calculated on a corrected correlation matrix of predictors and the criterion. Correlations in this latter matrix were corrected for criterion unreliability and multivariate range restriction. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

[37] Half the items in the E6 version of the SJT differ from those in the E4/E5 version. In addition, item selection for each version was based on their relation with the observed performance composite at the relevant pay grade.

For E5 Soldiers, most predictors showed significant incremental validity beyond SimPPW. All of the ExAct incremental validity estimates, however, were below .05. In most cases, there was less incremental validity for observed performance than for expected future performance. However, all of the prediction instruments, except the ExAct, added substantially to the prediction of observed and expected future performance.

For E6 Soldiers, the incremental validity estimates for most predictors were near zero when predicting observed performance; only the SJT and SJT-X had incremental validity estimates above .04. Similarly, when predicting future performance, only the SJT, SJT-X, ExAct Computer Experience, and AIM Adjustment scales had incremental validity estimates over .04.

Multiple Regression Analyses with All Predictors

Tables 9.7 and 9.8 show results of multiple regression analyses where the observed performance and expected future performance composites were used as the outcome variable, respectively. These analyses were conducted for purely theoretical purposes; there is no proposal to use all of the instruments together at the same promotion decision point.

With all predictors entered into the regression equation, the multiple *R* (correcting for unreliability in the criterion, range restriction in the predictors, and shrinkage) was very high for E5 Soldiers ($R = .50$ for observed performance, $R = .67$ for future performance) and moderate for E6 Soldiers ($R = .32$ for observed performance, $R = .37$ for future performance). Consistent with previous results in this chapter, the validity estimates were much lower for E6 Soldiers than for E5 Soldiers.

The relative contributions of the individual predictor scores to the prediction of performance were evaluated using dominance analysis (Johnson, 2001). The relative weights and the regression weights provided similar results. For example, for observed performance among E5 Soldiers, the top two scores were AIM Work Orientation and BIQ Leadership, which were followed by the SJT and the interview. For observed performance among E6 Soldiers, the SJT contributed the most to predicting performance. SimPPW and the ExAct General Experience score were next. The SJT-X, BIQ Interpersonal Skill, ASVAB GT, and BIQ Manipulativeness also had meaningful contributions to predicting performance for E6 Soldiers. Interestingly, for observed performance AIM Leadership did not contribute to prediction when the other predictors were included in the regression equations.

With all of the predictor scores in the regression equations, several regression coefficients became negative. AIM Agreeableness had the largest negative coefficients; ExAct Supervisory Experience also has large negative coefficients. It appears that many of these predictors act as suppressor variables. That is, they have little or no relation to the criterion, but they share significant variance with some predictors that *are* related to the criterion. This removal of non-predictive variance from predictors increases *R*. Some of this non-predictive variance is probably method variance. As a practical matter, a prediction battery is not likely to include variables with negative weights when, conceptually, these variables are positively related to the criterion. For example, candidates would not understand if they had points deducted because they have a lot of supervisory experience. Therefore, the multiple *R* values shown are higher than those that would be in obtained in practice.

Table 9.7. Regression of the Observed Performance Composite on All Predictor Scores by Pay Grade

Predictor	Raw				Corrected			
	β		Rel. Wt. (%)		β		Rel. Wt. (%)	
	E5	E6	E5	E6	E5	E6	E5	E6
SimPPW Composite	.10	.10	8.0	10.4	.10	.15	2.7	13.5
ASVAB GT Score	.01	.05	0.5	4.5	.02	.09	0.7	5.8
SJT Composite	**.14**	**.14**	15.1	17.0	.24	.21	15.5	18.0
SJT-X Composite	.	.08	.	10.5	.	.11	.	9.8
Interview Composite	**.11**	.	10.6	.	.16	.	12.0	.
ExAct Computer Experience	-.01	.05	1.3	3.6	-.01	.07	0.4	3.3
ExAct Supervisory Experience	-.05	-.09	0.8	3.2	-.13	-.13	1.7	4.3
ExAct General Experience	-.04	.13	1.5	10.3	-.05	.18	0.7	11.2
AIM Dependability	.01	-.11	0.8	7.3	.01	-.15	1.5	3.8
AIM Adjustment	-.04	.01	0.7	0.9	-.05	.01	0.8	0.5
AIM Work Orientation	**.20**	.04	19.7	1.4	.29	.05	18.7	1.2
AIM Agreeableness	-.08	-.11	1.5	6.4	-.11	-.15	1.4	4.6
AIM Physical Conditioning	.06	.02	4.3	0.2	.09	.02	7.2	0.4
AIM Leadership	-.06	-.03	6.1	0.5	-.09	-.05	5.6	0.7
BIQ Hostility to Authority	.05	-.04	0.6	5.0	.07	-.05	0.7	3.9
BIQ Manipulativeness	.08	-.09	1.0	4.8	.11	-.12	1.0	5.9
BIQ Social Perceptiveness	-.03	-.06	4.1	2.4	-.05	-.08	4.2	2.2
BIQ Social Maturity	.01	-.02	0.9	1.5	.02	-.03	0.5	1.5
BIQ Tolerance for Ambiguity	.06	-.06	3.1	0.7	.08	-.09	3.3	0.8
BIQ Openness	-.09	-.03	1.4	2.4	-.11	-.05	1.7	1.6
BIQ Leadership	**.23**	-.02	16.3	0.7	.30	-.02	17.8	1.0
BIQ Interpersonal Skill	.03	.10	1.8	6.5	.05	.13	2.0	6.0

Statistic	Overall Model Statistics			
	Raw		Corrected	
	E5	E6	E5	E6
R (all predictors)	**.40**	.33	.57	.47
ΔR (all predictors beyond SimPPW)	**.24**	.22	.44	.26
Shrunken R	.27	.00[a]	.50	.32
Δ Shrunken R	.11	.00	.38	.11

Note. $n_{E5} = 432$; $n_{E6} = 296$. [a] The shrunken R^2 value was estimated to be less than zero. "Raw" coefficients were calculated on the uncorrected correlation matrix of predictors and the criterion. "Corrected" coefficients were calculated on a corrected correlation matrix of predictors and the criterion. Correlations in this latter matrix were corrected for criterion unreliability and multivariate range restriction. "Rel. Wt." the relative weight of each predictor expressed in terms of the percentage of R^2 it accounts for relative to other predictors. "Shrunken" values represent observed multiple correlation values adjusted for shrinkage using Rozeboom's (1978) formula. Bolded values are statistically significant, $p < .05$ (one-tailed).

Table 9.8. Regression of the Expected Future Performance Composite on All Predictor Scores by Pay Grade

Predictor	Raw				Corrected			
	β		Rel. Wt. (%)		β		Rel. Wt. (%)	
	E5	E6	E5	E6	E5	E6	E5	E6
SimPPW Composite	.01	.06	0.7	6.2	.01	.09	0.3	6.7
ASVAB GT Score	-.02	.03	0.1	3.1	-.03	.06	0.2	3.0
SJT Composite	**.13**	.09	10.8	7.4	.25	.16	10.6	11.1
SJT-X Composite	.	.09	.	12.7	.	.14	.	11.6
Interview Composite	.08	.	6.8	.	.14	.	8.2	.
ExAct Computer Experience	.00	.12	1.0	15.1	.00	.20	0.4	16.4
ExAct Supervisory Experience	-.05	-.02	0.8	0.4	-.14	-.04	0.9	0.6
ExAct General Experience	-.03	.06	1.1	6.6	-.04	.11	0.5	6.2
AIM Dependability	.10	-.07	3.0	4.2	.17	-.10	4.2	2.2
AIM Adjustment	-.08	.15	1.6	9.4	-.11	.23	1.3	12.7
AIM Work Orientation	**.23**	.07	23.2	2.9	.36	.10	21.7	2.5
AIM Agreeableness	**-.16**	**-.17**	5.6	12.0	-.25	-.25	3.9	9.1
AIM Physical Conditioning	.06	.00	3.5	0.6	.10	.00	7.1	0.5
AIM Leadership	-.01	-.05	10.2	0.6	-.01	-.08	8.3	0.8
BIQ Hostility to Authority	-.08	.02	1.8	1.4	-.12	.03	2.2	0.9
BIQ Manipulativeness	.04	-.03	1.3	1.8	.07	-.05	1.3	2.2
BIQ Social Perceptiveness	-.09	.05	3.7	1.4	-.13	.07	3.3	1.1
BIQ Social Maturity	-.13	.01	1.7	0.7	-.22	.01	2.6	0.7
BIQ Tolerance for Ambiguity	-.02	-.02	0.8	0.7	-.02	-.03	0.7	0.7
BIQ Openness	-.04	-.11	1.0	7.2	-.06	-.18	0.7	6.1
BIQ Leadership	**.26**	-.04	19.3	0.7	.40	-.07	19.3	0.8
BIQ Interpersonal Skill	.04	.08	1.9	5.0	.06	.12	2.2	4.2

Statistic	Overall Model Statistics			
	Raw		Corrected	
	E5	E6	E5	E6
R (all predictors)	**.43**	.31	.70	.50
Δ R (all predictors beyond SimPPW)	**.35**	.21	.64	.32
Shrunken *R*	.32	.00[a]	.67	.37
Δ Shrunken *R*	.24	.00	.60	.19

Note. $n_{E5} = 435$; $n_{E6} = 300$. [a] The shrunken R^2 value was estimated to be less than zero. "Raw" coefficients were calculated on the uncorrected correlation matrix of predictors and the criterion. "Corrected" coefficients were calculated on a corrected correlation matrix of predictors and the criterion. Correlations in this latter matrix were corrected for criterion unreliability and multivariate range restriction. "Rel. Wt." the relative weight of each predictor expressed in terms of the percentage of R^2 it accounts for relative to other predictors. "Shrunken" values represent observed multiple correlation values adjusted for shrinkage using Rozeboom's (1978) formula. Bolded values are statistically significant, $p < .05$ (one-tailed).

These analyses show that, when a predictor battery is put together, each predictor must be considered in combination with other predictors rather than just by itself. In addition the results of the dominance analysis are conditional on the predictors entered into the battery. Changing even one predictor (by addition or deletion) could dramatically alter the results. The target pay grade must also be considered. Given these caveats, the following scales appeared to performed well for E5 Soldiers regardless of the other predictors for both observed and expected future performance: SimPPW, SJT, Interview, AIM Work Orientation, and BIQ Leadership. For E6 Soldiers, the following predictors did consistently well for both criteria: SimPPW, SJT, SJT-X, ExAct General Experience, and BIQ Interpersonal Skill.

The best set of predictors depends, to some degree, on whether the criterion is observed performance or expected future performance. The ASVAB GT was slightly less predictive and AIM Work Orientation was more predictive of expected future performance (compared with observed performance) for both E5 and E6 Soldiers. Among only E5 Soldiers, AIM Dependability and BIQ Leadership were more predictive of future performance. Among only E6 Soldiers, ExAct Computer Experience and AIM Adjustment were considerably more predictive of future performance. Thus, the trend is that personality attributes were slightly more predictive and general cognitive ability was slightly less predictive of future performance (compared with observed performance).

Summary: Criterion-Related Validity

The validity analyses—zero-order correlations and incremental validity estimates—provide similar results for the individual predictor measures. All of the predictor measures yield one or more scores that show validity evidence, though some scores were more effective than others in yielding incremental validity over the simulated PPW score. Clearly, the findings differ for E5 Soldiers versus E6 Soldiers. As discussed in the next section, the results also varied across CMF.

Additional Validation Analysis Issues

During the course of our validity analysis work, one observation repeatedly surfaced. This was that there were differences in the size and pattern of criterion-related validity across pay grades. As we explored reasons for these differences, we discovered that such differences were also observed across Soldiers from different types of MOS. We close this chapter with a closer examination of these two findings.

Validity Differences between E5 and E6 Soldiers

We examined several hypotheses regarding the source of differences in the criterion-related validity evidence observed across pay grades. Unfortunately, with one exception, the data supported none of our hypotheses (summarized in Table 9.9). What remains is the possibility that the instruments we examined are simply a better match to our E5 than our E6 Soldier performance criteria.

The one exception is related to the potential differential functioning of the SJT across pay grades. Analyzing data from this project, Putka, Waugh, and Knapp (2002) showed that tenure within pay grade had a moderating effect on the relationship between SJT and observed performance composite scores when all 40 SJT items were scored for E6 Soldiers. They showed evidence of a disordinal interaction between time in grade and SJT scores when predicting

observed performance. Specifically, among E6 Soldiers with low time in grade, SJT scores show a strong *positive* correlation with observed performance; in contrast, among E6 Soldiers with high time in grade, the correlation is *negative*. Thus, inclusion of E6 Soldiers high in tenure may have resulted in attenuated E6 validity estimates (relative to E5 validity estimates).

Validity Differences for Soldiers in Different CMF

While exploring potential explanations for E5-E6 Soldier validity estimate differences, we considered the possibility of compositional differences in the E5-E6 Soldier samples in terms of the CMF membership. Although the E5 and E6 Soldier samples comprised similar proportions of Soldiers in CMF, there were several sizable differences in the criterion-related validity estimates for predictor scores across CMF. Given such findings, and because the current promotion system is currently uniform across CMF, we further pursued potential differences in validity by CMF for each predictor.

Table 9.9. Hypothesized Explanations for Observed E5-E6 Validity Differences

Statistical Artifacts
• Greater range restriction on criteria for E6 Soldiers relative to E5 Soldiers
• Lower internal consistency among ratings on scales forming the E6 criterion composites relative to the E5 criterion composites
• Preponderance of influential data points that negatively affect E6 correlations, or unduly positively affect E5 correlations
• Non-linearity in the relationship between predictors and criteria for E6 Soldiers but not E5 Soldiers (Pearson correlations do not fully account for non-linear relationships)
• Differences in the amount of intrarater variance for E5 and E6 Soldiers (indicator of halo tendency)
Differences in the Meaning of Job Performance across E5 and E6 Samples
• Differences between E5 and E6 samples in the predictiveness of each dimension-specific rating scale when overall effectiveness was used as the criterion (e.g., policy capturing analysis)
• Differential rank-ordering of the variance of ratings on scales forming criterion composites for E5 and E6 Soldiers such that the rating scales that are most easily predicted have less variance for E6 Soldiers than E5 Soldiers, and ratings scales that are less easily predicted have more variance for E6 Soldiers than E5 Soldiers (unit-weighted criterion composites effectively give scales with more variance greater weight)
• Differences in rater confidence for expected future performance ratings (perhaps future E6 performance is more difficult to predict)
Substantive Differences between E5 and E6 Samples
• Composition differences of E5 and E6 Soldier samples in terms of race, gender, CMF category, length of rater supervision, distance between supervision pay grades of supervisors and Soldiers rated, proportion of mail-backs (where such composition variables covary with the criteria)
• Differential moderating effects of "tenure in pay grade" for E5s and E6s (E6s have greater range of tenure in pay grade. To the extent that predictor validities drop off at higher levels of tenure within pay grade, E6 zero-order validities may be attenuated relative to E5 zero-order validities.)

Unfortunately, there were not enough Soldiers in our sample from each CMF to investigate this issue thoroughly. Therefore, we explored the potential for differential prediction by CMF by focusing on CMF *categories* that had sufficient sample sizes ($n > 100$) for relatively

stable validity estimates to emerge. Based on this criterion, we were able to compare validity estimates for two of the six CMF categories (Combat Operations and Logistics). The estimated validities (i.e., zero-order correlations with the criteria) of each predictor within these two CMF categories, broken down by pay grade and type of criterion (observed performance vs. expected future performance), are presented in Table 9.10.

The correlations presented in Table 9.10 indicate that the SimPPW Composite was a significantly better predictor of E5 Logistics Soldiers' than of E5 Combat Operations Soldiers' performance (both observed and future). The lack of other significant differences sufficient to show an interpretable pattern may be due to the relatively small sample sizes.

Table 9.10. Raw Correlations between Predictor and Criterion Scores for Soldiers in Combat Operations and Logistics CMF Categories (by Pay Grade)

Predictor	Observed Performance Composite				Expected Future Performance Composite			
	E5		E6		E5		E6	
	Com	Log	Com	Log	Com	Log	Com	Log
ASVAB GT Score	.05	.02	.12	.12	.08	-.03	**.15**	.13
SJT Composite	**.28**	**.26**	**.19**	.09	**.31**	**.16**	**.20**	.02
SJT-X Composite	.	.	**.22**	.15	.	.	**.19**	**.20**
Interview Composite	**.22**	**.20**	.	.	**.22**	**.18**	.	.
SimPPW Composite	$.01_a$	**.35**	.08	-.04	$-.01_a$	**.24**	.10	.08
ExAct Computer Experience	**.12**	.11	**.17**	-.06	**.17**	.06	**.23**	.05
ExAct Supervisory Experience	**.11**	.11	-.08	-.07	**.12**	**.13**	-.05	-.01
ExAct General Experience	**.15**	**.17**	.09	.12	**.14**	**.21**	.11	.08
AIM Dependability Scale	.05	.11	$\mathbf{-.22_a}$	.08	**.11**	.09	-.13	.07
AIM Adjustment Scale	**.14**	-.01	.04	-.03	**.11**	-.04	.11	.01
AIM Work Orientation Scale	**.25**	**.31**	-.01	.13	**.31**	**.26**	.06	.12
AIM Agreeableness Scale	.06	.02	$-.16_a$	**.18**	.04	-.09	-.07	.13
AIM Physical Conditioning Scale	**.15**	**.14**	.08	-.08	**.17**	**.15**	$.11_a$	-.14
AIM Leadership Scale	**.18**	**.26**	.06	-.02	**.27**	**.27**	.08	.02
BIQ Hostility to Authority	.02	-.10	-.07	.01	-.03	-.08	-.05	-.02
BIQ Manipulativeness	.01	-.09	-.07	.05	-.06	-.04	-.06	-.03
BIQ Social Perceptiveness	**.21**	**.12**	.01	.01	**.26**	.09	.06	.05
BIQ Social Maturity	-.06	.09	.04	.01	-.07	.04	.07	.06
BIQ Tolerance for Ambiguity	.10	**.17**	.04	.06	.10	**.17**	.11	.09
BIQ Openness	**.13**	-.04	-.01	-.09	.15	.02	.01	-.06
BIQ Leadership	**.25**	**.24**	.10	-.05	**.33**	**.25**	**.14**	.00
BIQ Interpersonal Skill	.06	**.15**	**.15**	.05	.10	.10	**.19**	.10

Note. Com = Combat Ops, Log = Logistics. Combat Ops n_{E5} = 187-235; Combat Ops. n_{E6} = 135-158; Logistics n_{E5} = 175-208; Logistics n_{E6} = 106-123. The "a" subscripts on Combat Operations correlations indicate that corresponding Combat Operations and Logistics correlations (for the same pay grade and criterion) were significantly different from each other, $p < .05$ (two-tailed). Statistically significant correlations are bolded, $p < .05$ (one-tailed).

Other predictors had sizable (only a couple of which are significant) differences in the opposite direction. In at least two of four comparisons, Combat Operations Soldiers' validity estimates were substantially higher than those of Logistics Soldiers for the following predictors: SJT, ExAct Computer Experience, AIM Adjustment, BIQ Social Perceptiveness (E5 Soldiers only), BIQ Openness (E5 Soldiers only), and BIQ Interpersonal Skills (E6 Soldiers only). Finally, AIM Dependability and Agreeableness correlated negatively with performance for E6 Combat Operations Soldiers and positively with performance for Logistics Soldiers. The relative importance of KSAs differ between these jobs. Therefore, it is not surprising that predictors' validity estimates differ somewhat between the two different CMF.

Although some CMF-based validity differences were found, generalizing these results to other CMF categories should be done cautiously. Perhaps differences (or lack thereof) in validity found between Combat Operations and Logistics CMF categories would be less substantial, more substantial, or roughly similar across other specific CMF. Unfortunately, due to small sample sizes, it was not possible to evaluate these possibilities for other CMF categories with the current data.

Further research might be useful for understanding both the pay grade and CMF/CMF category differences in the criterion-related validity of the various NCO21 predictors. The analyses presented in this report are based on sufficiently large samples, however, to point clearly to the conclusion that the differences are real.

Summary

Examining the relations among the predictors and criteria yielded some noteworthy results. The observed relations among the predictor scores generally support their construct validity. Overall, the examined predictor scores showed a level of incremental validity such that they could substantially improve the E4-to-E5 and E5-to-E6 Soldier promotion system. In additional, other findings suggest further investigation of the following: (a) individual differences on personality/temperament constructs seem to have different relations with E5 and E6 Soldier judgment and performance, (b) the NCO21 predictors predict E5 Soldier performance better than they predict E6 Soldier performance, and (c) some predictors might correlate more highly with performance in some MOS than in others.

CHAPTER 10: SUMMARY

Deirdre J. Knapp and John P. Campbell
HumRRO

As described in Chapter 1, the goal of the NCO21 project is to help the Army understand and plan for the impact of future performance demands on the NCO performance management system. Particular attention has been given to the semi-centralized promotion system, but the information and tools derived from this research may also support improvements to training and development activities.

Early stages of the NCO21 project produced future-oriented job analysis information that was used as a basis for identifying and developing predictor and job performance criterion measures. The predictors included a situational judgment test, semi-structured interview, self-report record of experience, two temperament inventories, and the ASVAB. They also included a self-report form to collect information used to calculate Promotion Point Worksheet points according to the current semi-centralized promotion system. The criterion measures were two supervisor rating instruments, one pertaining to current performance and the other pertaining to expected performance under future Army conditions.

In this last stage of the project, we administered the predictor and criterion measures to a sample of Soldiers across a variety of MOS and locations. The purpose of the present report has been to document the analyses of these data as they relate to the psychometric properties and validity of the NCO21 measures.

Empirical Results

Overall, the results of the validation analyses were very promising. All of the predictor instruments yielded one or more scores that were significantly correlated with performance, both current and future. Even when examining incremental validity over the current system, most instruments held their own. Complicating the analyses and subsequent conclusions was the finding that the empirical results varied across pay grade and CMF. Despite extensive analyses to identify artifactual source(s) of these differences (e.g., range restriction), none were found.

Important Caveats

It is important to bear in mind certain limitations to the NCO21 research design when interpreting the empirical findings. We will discuss several here, including the (a) limited scope of the criterion measures, (b) concurrent nature of the design, and (c) limitations of generalizability to an operational context. Although not a limit of the research design, another important caveat pertains to the limited scope of the analyses we conducted using the self-report PPW information.

Criterion Measurement

Although the two rating instruments used in the NCO21 research had broad coverage, prior research has shown that measurement method can make a big difference in observed criterion scores. For example, in the Army's Project A, ratings of MOS technical knowledge and skill were not highly

correlated with more direct measures of this performance area (i.e., written multiple-choice and hands-on work sample tests) (J. Campbell & Knapp, 2001). Rather, ratings were most useful for assessing "will-do" aspects of performance but greater confidence was given to the written and hands-on tests for assessing "can-do" aspects of performance. The wide array of predictor measures also showed distinct patterns in which some scores (e.g., from the temperament inventory) predicted will-do performance well, but others (e.g., ASVAB subtest scores) predicted can-do performance.

It was beyond the scope of the NCO21 project to develop and administer performance tests and there were no operational scores of record (such as the old Skill Qualification Test scores) that could be used. Therefore, it is quite possible that some NCO21 predictors would look more or less attractive if we had evidence of their validity for predicting "can-do" performance at the E5 and E6 pay grades. Results related to the Aptima-developed computer simulation (to be reported separately) may provide some related evidence, but it will be quite limited because of small sample sizes.

Concurrent Design

The concurrent design of the NCO21 project enabled the research to be conducted in a relatively short timeframe. It is also reasonable to believe that a predictor that demonstrates criterion-related validity in a concurrent setting is likely to demonstrate validity in a longitudinal setting. What is less convincing, however, is the accuracy with which we can estimate the best ways to combine or weight scores from different measures to produce the most effective promotion decisions using concurrent data. The problem is particularly acute here because it is reasonable to speculate that performance on several of the predictor measures used in NCO21 (the SJT, interview, SimPPW, and ExAct) is influenced by experience and training. Indeed, it may well be that these measures would yield even higher criterion-related validity in a longitudinal setting. In any case, the validity and optimal weighting of the various NCO21 predictors should be examined in a longitudinal setting.

A related observation is that limited resources (time and personnel) prevented administration of the NCO21 interview to Soldiers in all three target pay grades (E4, E5, E6). We wanted to make sure the interview was suitable for E4 Soldiers seeking promotion to E5, but this meant that the interview could not be administered to E6 Soldiers. Although the interview was not developed for E6 Soldiers, not having interview data for them in a concurrent validation meant we would be unable to evaluate the validity of the interview for predicting E6 performance. Relevant data could be collected in a longitudinal study.

Research vs. Operational Context

The research setting is an inherently imperfect reflection of operational conditions. Of particular concern is the motivation of the participants. In a research setting, participants do not have a strong vested interest in their performance. We can encourage them to do their best on the measures in the interests of the goals of the research, but this is not the same as knowing their performance will determine their qualifications for promotion. Indeed, in an operational setting, the motivation to perform well can lead to efforts to beat the system by cheating on tests (e.g., memorizing a scoring key) or faking on self-report inventories (e.g., endorsing all the leadership-related items on a temperament measure).

Testing professionals have many strategies for addressing the problem of test compromise in operational settings. These include security measures and multiple test forms. Though imperfect, such strategies are generally effective when dealing with maximal performance tests that involve the assessment of abilities. The testing community has been less successful at handling the phenomenon of faking temperament measures. AIM uses one well-known method—a forced-choice item format. But the first large-scale use of the AIM in an operational setting (the Army's pilot GED Plus program) showed criterion-related validity far below that anticipated based on research findings, presumably due to Army applicants "faking good" on the measure (Knapp, Waters, & Heggestad, 2002).

While we are particularly concerned about the generalizability of our findings to an operational setting for the AIM and BIQ, it is possible that some of the other NCO21 predictors will also perform somewhat differently in an operational setting. At a minimum, any measure adopted for operational use in the Army's semi-centralized promotion system will need to address concerns related to compromise. For example, there is relatively little literature related to the development of parallel SJT forms, but this will certainly be a requirement for implementation in the Army.

Optimization of PPW Information

The Promotion Point Worksheet contains dozens of items that we combined and scored to be as consistent as possible with how the Army currently assigns promotion points (with the important limitation that we had no way to simulate board or Commander's points). However, there are an almost infinite number of ways this information could be scored, some of which would likely improve the criterion-related validity of the instrument. We could, for example, investigate different ways of computing the four administrative PPW subscores (i.e., Awards, Military Education, Civilian Education, Military Training), such as removing the point limits currently imposed or giving different numbers of points for various accomplishments. Consider the PPW Awards score. There are over two dozen individual awards, each of which is worth from 3 to 35 points. Conceivably, we could conduct analyses (and gather input from Army SMEs) that would suggest different point allocations for each award.

The point to be made here is simply that the Army would likely benefit from simple scoring changes in the current Promotion Point Worksheet, without the addition of any new predictor measures. Analyses to support such changes were not reported here, in part because our focus was on the incremental validity of the experimental measures over the current system and because it would be preferable to conduct such analyses on longitudinal data.

Next Steps

This report has focused on the NCO21 project's empirical validation findings, whereas there are policy concerns, practical considerations, and findings from additional research that would need to factor into any specific implementation decisions. A companion report (Knapp & Heffner, 2003) discusses implementation-related issues, ideas, and recommendations that build on the empirical results reported here.

REFERENCES

Borman, W. C., Motowidlo, S. J., Rose, S. R., & Hanser, L. M. (1985). *Development of a model of soldier effectiveness* (ARI Technical Report 741). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Campbell, J. P., & Knapp, D. J. (2001). *Exploring the Limits in Personnel Selection and Classification*. Mahwah, NJ: Lawrence Erlbaum Associates.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43,* 313-333.

Campbell, J. P., Walker, C. B., & Knapp, D. J. (1998). *Soldier Characteristics for the 21st Century Part I: Methodology* (ARI Report). Alexandria, VA U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco, CA: Jossey-Bass.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement, 5*, 115-124.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Stamford, CT: Thomson Learning.

Ford, L. A., Campbell, R. C., Campbell, J. P., Knapp, D. J, & Walker, C. B. (2000). *21st Century Soldiers and Noncommissioned Officers: Critical Predictors of Performance* (ARI Technical Report 1102). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hess, K., Entin, E., Garrity, M. Baker, K. Miller, P., Baker, B., Putka, D. J., & Lorenzen, C. (2002). *Simulation testbed for evaluating performance (STEP): Final Report* (Contract No. DASW01-00-C-3012). Woburn, MA: Aptima Human-Centered Engineering.

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74,* 167-184.

Johnson, J. W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgments of overall performance. *Journal of Applied Psychology, 86*, 984-996.

Kilcullen, R. N., Chen, G., Zazanis, M. M., Carpenter, T., & Goodwin, G. (1999, April). *Adaptable performance in unstructured environments.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Kilcullen, R. N., Mael, F. A., Goodwin, G. F., & Zazanis, M. M. (1999). Predicting U.S. Army Special Forces Field Performance. *Journal of Human Performance in Extreme Environments, 4*, 53-63.

Kilcullen, R. N., White, L. A., Mumford, M. D., & Mack, H. (1995). Assessing the construct validity of rational biodata scales. *Military Psychology, 7*, 17-28.

Kilcullen, R. N., White, L. A., Zacarro, S., & Parker, C. (2000, April*). Predicting managerial and executive performance.* Paper presented at the 15th annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Knapp D. J., Heffner, T. S., & Campbell, R.C. (2003). *Recommendations for an Army NCO semi-centralized promotion system for the 21st Century.* (ARI Research Report 18-07). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., Burnfield, J. L, Sager, C. E., Waugh, G. W., Campbell, J. P., Reeve, C. L., Campbell, R. C., White, L. A., & Heffner, T. S. (2002). *Development of predictor and criterion measures for the NCO21 research program* (Technical Report 1128). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., Campbell, C. H., Borman, W. C., Pulakos, E. D., & Hanson, M. A. (2001). Performance assessment for a population of jobs. In J. P. Campbell and D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 181-235). Mahwah, NJ: Lawrence Erlbaum Associates.

Knapp, D. J., Waters, B. K., & Heggestad, E. D. (2002). *Investigations related to the implementation of the Assessment of Individual Motivation* (Confidential Interim Report). Alexandria, VA: Human Resources Research Organization.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*, 493-505.

McGraw, K. O. & Wong, S. P. (1996). Forming inferences about some intraclass correlations. *Psychological Methods, 1*, 30-46.

Nunnally, J. C. (1978). *Psychometric Theory.* New York: McGraw-Hill.

Putka, D. J., Waugh, G. W., & Knapp, D. J. (2002, October). Pay grade differences in the functioning of the situational judgment test. In G. W. Waugh (Chair) *Tailoring a Situational Judgment Test to Different Pay Grades*. Symposium presented at the 44th Annual Conference of the International Military Testing Association, Ottawa.

Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlation: A clarification. *Psychological Bulletin, 85*, 1348-1351.

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relationships between measures of typical and maximum job performance. *Journal of Applied Psychology, 37,* 482-486.

Stanton, E. F., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and Strategies for Reducing the Length of Self-Report Scales. *Personnel Psychology, 55*, 167–194.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement technique.* New York: Wiley.

Viswesvaran, C., Ones, D.S., & Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557-574.

White, L. A., (2002, October). *A Quasi-Ipsative Temperament Measure for Predicting Future NCO Leadership Performance.* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

White, L. A., & Young, M. C. (1998, August). *Development and validation of the Assessment of Individual Motivation (AIM).* Paper presented at the annual meeting of the American Psychological Association, Washington, DC

White, L. A., & Young, M. C. (2001, April). *Validation of a faking-resistance measure of temperament constructs.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

Young, M. C., Heggestad, E. D., Rumsey, M. G., & White, L. A. (2000, August). *Army Pre-implementation Research Findings on the Assessment of Individual Motivation (AIM).* Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Appendix A
Observed Performance Rating Scales

Section I: Observed Performance Rating Scales

1. MOS/Occupation-Specific Knowledge and Skill						
How effectively does this soldier display job-specific knowledge and skill?						
Does not display the knowledge or skill required to perform many work assignments or tasks; is unaware of recent developments relevant to his/her MOS.		Displays adequate knowledge of most aspects of the job; has sufficient skills to handle moderately difficult problems and to get most assignments done properly; attempts to keep informed of most important developments in his/her MOS.			Is highly competent in performing the technical tasks for which he/she is responsible; has skills and technical knowledge necessary to handle difficult problems; strives to stay informed of latest developments in his/her MOS.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

2. Common Task Knowledge and Skill						
How effectively does this soldier display the necessary knowledge and skill to perform common tasks?						
Does not display the knowledge or skill required to perform common assignments or tasks (e.g., land navigation, field survival techniques, NBC protection).		Displays good knowledge of most common areas; has sufficient skills to handle moderately difficult problems and to perform common tasks properly.			Is highly competent in performing common tasks; possesses skills and knowledge necessary to handle most common tasks, even under difficult conditions.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

3. Computer Skills						
To what extent does this soldier display an understanding of computer systems, operating systems, and applications?						
Does not display any understanding of computers above basic usage or Windows-based applications; cannot troubleshoot even the most basic application errors.		Displays basic understanding of some operating systems (e.g., DOS, Windows NT); can troubleshoot basic application errors; can troubleshoot simple systems errors; understands computer terminology.			Is highly competent administrating most operating systems (e.g., DOS, Windows NT, Army specific); can troubleshoot serious application errors; can set up and troubleshoot computer systems; well versed in computer terminology.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

4. Writing Skill						
How effectively does this soldier prepare written materials?						
Usually writes in an awkward or confusing manner; uses incorrect grammar, punctuation, and spelling; often includes irrelevant information in the material; written products often require a lot of editing.		Typically writes logically but will occasionally make grammatical, punctuation, or spelling errors; usually includes most relevant information and tries to tailor the work to the audience; written products sometimes require editing.			Usually writes concisely, clearly, and logically; focuses on relevant issues; uses correct grammar, punctuation, and spelling; effectively tailors the work to the audience; written products require little or no editing.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

5. Oral Communication Skill						
How effectively does this soldier orally communicate?						
Speaks in an awkward or confusing manner; does not present ideas clearly; often rambles or strays to irrelevant topics; mispronounces words or terms; speaks too fast or too slow.		Usually expresses him or herself clearly and logically; makes few grammatical errors; typically gets information across effectively; generally speaks at an appropriate, smooth pace.			Always expresses him or herself clearly and logically; gets to the point quickly; uses correct grammar; appropriately tailors the presentation to the audience; focuses on relevant and important issues; always speaks fluently and at a smooth pace.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

6. Level of Effort and Initiative on the Job						
To what extent does this soldier put forth effort and initiative on the job/mission/assignment?						
Shows little effort or initiative to accomplish tasks; completes assignments carelessly; often fails to meet deadlines; rarely seeks out additional responsibilities or challenging tasks.		Demonstrates sufficient effort on most tasks and assignments; is usually reliable about completing assignments on time; puts forth extra effort when necessary; sometimes seeks out additional responsibilities, training, or challenging tasks.			Shows a lot of initiative and often puts forth extra effort to get tasks done effectively, even under difficult conditions; reliably accomplishes work on time; enthusiastically takes on challenging assignments and additional responsibilities.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

7. Adaptability		
How effectively does this soldier adapt to varying environments by modifying behavior, plans, or goals?		
Has difficulty functioning effectively in new situations; does not adapt quickly to new environments, people, or equipment; is easily frustrated in situations that do not go as planned.	Is able to function adequately in new situations; modifies behavior when faced with unexpected events or conditions; adapts fairly readily to new people, situations, or equipment.	Thinks and acts quickly in response to changes in the environment; often develops innovative and imaginative approaches to dealing with unexpected events; can effectively change plans when the situation requires it.
LOW	**MODERATE**	**HIGH**
1 2	**3 4 5**	**6 7**

8. Self-Management and Self-Directed Learning Skill		
How effectively does this soldier self-manage his/her job responsibilities, training and career development, and personal responsibilities?		
Makes little or no effort to balance work and personal responsibilities; uses finances irresponsibly; ignores or otherwise fails to participate in relevant career training opportunities; needs constant supervision; fails to seek advice when needed.	Shows effort to manage work and personal responsibilities; typically uses finances responsibly; participates in required courses/training; attempts to work on problem areas when encouraged to do so; can usually work independently; seeks advice when needed but sometimes from inappropriate sources.	Effectively manages work and personal responsibilities; demonstrates exceptional financial responsibility; studies and works hard during off-duty hours to improve job-related skills; actively seeks additional responsibilities to improve job skills and increase chance of promotion; works well without supervision; willingly seeks advice when appropriate.
LOW	**MODERATE**	**HIGH**
1 2	**3 4 5**	**6 7**

9. Demonstrated Integrity, Discipline, and Adherence to Army Procedures		
To what extent does this soldier adhere to Army procedures and values, and demonstrate integrity, ethical behavior, and self-discipline on the job?		
Is disrespectful toward superiors; is sometimes dishonest; has difficulty accepting and following superiors' orders; makes up excuses to avoid assignments; fails to take responsibility for his/her job-related errors; often fails to follow rules, policies, and regulations; takes unnecessary risks that endanger the safety of self and/or others.	Is usually respectful to superiors; is generally honest; obeys direct orders; takes responsibility for most job-related mistakes he/she makes; usually attempts to follow applicable rules, policies, and regulations; typically avoids unnecessary risks and notices potential safety hazards.	Is always respectful to superiors; is honest about work matters, even when it may go against personal interests; obeys orders; ensures others are not blamed for his/her mistakes; carefully follows rules, policies, and regulations; tries to make sure others follow the rules; takes steps to protect self and others from safety risks.
LOW	**MODERATE**	**HIGH**
1 2	**3 4 5**	**6 7**

10. Acting as a Role Model						
To what extent does this soldier set a good example for others to follow in terms of physical fitness, military bearing, and appropriate behavior?						
Is generally overweight or in poor physical condition; avoids exercise; often dresses sloppily; displays poor military bearing; sets a poor example for others to follow and fails to model even minimally acceptable behavior as a soldier.		Meets basic standards for physical fitness; dresses properly, maintaining Army standards; usually displays good military bearing; attempts to set a good example of soldier behavior for others to follow.			Exercises consistently to maintain excellent physical fitness; always dresses sharply in correct uniform; consistently maintains excellent military bearing; sets an outstandin example for others by exceeding the standard for appropriate military behavior.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

11. Relating to and Supporting Peers						
How effectively does this soldier relate to and support peers?						
Tends to be rude, selfish, and insensitive to peers' concerns; generally fails to provide assistance to others, even when there is a clear need to do so; may force his/her approach to tasks on others without seeking input.		Usually courteous and tactful when dealing with peers; provides assistance to others, especially when it is clear that help is needed; tries to develop approaches to tasks that take into account obvious differences of opinion.			Always treats peers in a courteous and tactful manner; offers assistance without waiting to be asked, even in situations that involve complicated interpersonal situations; actively seeks out peers' opinions and incorporates peers' ideas into own plans.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

12. Cultural Tolerance						
How effectively does this soldier demonstrate tolerance and understanding of other cultural and social backgrounds both in the context of the diversity of U.S. Army personnel and interactions with foreign nationals?						
Does not understand or show respect for other cultural practices or beliefs; makes insensitive comments or slurs to others based on social or cultural differences, (e.g., racial heritage, religious beliefs, ethnic customs, language); cannot work, socialize, or communicate effectively with others from different backgrounds.		Recognizes need to be tolerant and respectful of other cultural, ethnic, and belief systems but does not always demonstrate understanding of social and cultural diversity; willing to work, communicate, and perhaps socialize with others from different backgrounds but does not do so easily.			Shows tolerance, understanding, and respect for other cultural, ethnic, and belief systems; shows respect for social and cultural diversity, (e.g., racial heritage, religious beliefs, ethnic customs, language); easily works, socializes, and communicates well with others regardless of differences in background.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

13. Selfless Service Orientation						
To what extent does this soldier display a selfless service orientation?						
Fails to support team or group; has a "looking out for number one" attitude; explicitly asks for credit for unselfish behavior.		Supports team or group when called upon to do so, but usually waits until asked; puts group or team goals ahead of own goals when it is easy to do so.			Willingly commits to the greater good of the team; willingly puts group or team goals ahead of individual goals when appropriate; does not expect credit for unselfish behavior.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

14. Leadership Skills						
To what extent does this soldier demonstrate strong leadership skills by effectively motivating, supporting and supervising individuals and being an effective team leader?						
Fails to support subordinates; does not reward effective behavior or provide useful feedback to improve performance; assigns duties unfairly; rarely makes sure assignments are understood and completed; does not communicate team goals; fails to lead team to adapt to mission changes; fails to resolve conflicts or does so unfairly.		Usually supports subordinates and rewards effective behavior; provides feedback to improve performance, but it is not always helpful; generally assigns work fairly; typically makes sure subordinates' work meets standards; communicates team goals but not always clearly; leads team to adapt to mission changes but takes time/effort to do so; attempts to resolve conflicts fairly.			Always supports subordinates and rewards effective behavior; maintains high morale; provides helpful feedback to improve performance; always assigns work fairly; always makes sure subordinates' assignments are understood and completed; clearly communicates team goals; leads team to adapt quickly to mission changes; resolves conflicts among subordinates fairly.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

15. Concern for Soldier Quality of Life						
How effectively does this soldier show consideration for subordinates' quality of life?						
Generally ignores subordinates' personal needs, constraints, and values; ignores or is insensitive to potential conflicts between subordinates' personal needs and duty demands; fails to show concern for the well-being of subordinates' personal lives.		Usually is aware of and attempts to help resolve conflicts between subordinates' work and personal needs; is sometimes sensitive to potential work/personal conflicts and attempts to help subordinates avoid such situations; shows basic awareness of subordinates personal needs, constraints, and values.			Has keen awareness of subordinates personal needs, constraints, and values; takes extra steps to resolve and avoid subordinate work/personal life conflicts; shows genuine concern for the well-being of subordinates' personal lives.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

16. Training Others		
How effectively does this soldier provide relevant training experiences for subordinates?		
Is unaware of or ignores individual or unit training needs; fails to provide training experiences or gives subordinates inappropriate training; does not prepare well for formal training situations; fails to guide subordinates on technical training matters.	Usually ensures that important subordinate training needs are met when made aware of such needs; uses existing classroom or on-the-job training techniques; prepares as required for training sessions; sometimes guides and tutors subordinates on technical matters.	Actively seeks to be aware of individual o unit training needs; always makes time to provide relevant formal and informal training experiences for subordinates; prepares thoroughly for training sessions; effectively guides and tutors subordinates on technical matters.
LOW	**MODERATE**	**HIGH**
1 2	**3 4 5**	**6 7**

17. Coordination of Multiple Units and Battlefield Functions		
To what extent does this soldier demonstrate knowledge of the interrelatedness among different units (including his/her own unit), as well as how to coordinate multiple battlefield functions?		
Cannot apply or coordinate multiple battlefield functions such as direct/indirect fires, communications, intelligence, and combat service support (CSS) to achieve tactical goals; shows little or no ability to understand how one unit's actions can affect the performance of other units; does not see how his/her unit's operations relate to the overall system.	Can apply and coordinate multiple battlefield functions (e.g., direct/indirect fires, communications, intelligence, CSS) with assistance; usually recognizes how one unit's actions can affect the performance of other units; understands how some goals and operations of own unit and other units relate but has difficulty analyzing the overall system.	Can independently apply and coordinate multiple battlefield functions (e.g., direct/indirect fires, communications, intelligence, and CSS) to achieve tactical goals; clearly understands how one unit's actions can affect the performance of other units; can quickly and accurately analyze how goals and operations of own unit relate to the overall system.
LOW	**MODERATE**	**HIGH**
1 2	**3 4 5**	**6 7**

18. Problem-Solving/Decision Making Skill		
How effectively does this soldier react to new problem situations and make reasonable, informed decisions regarding solutions?		
Usually reacts to new problem situations with frustration and confusion; fails to apply previous experience and training or realize their relevance; blindly applies rules or strategies without regard to the uniqueness of the situation; fails to assess costs or benefits of alternative solutions before making decisions.	Often reacts to new problem situations by applying previous experience or education/training, but does not always do so effectively; seldom applies rules or strategies blindly; attempts to assess costs and benefits of alternative solutions but does not always make timely decisions; has trouble making appropriate decisions with incomplete information.	Consistently reacts to new problem situations by applying previous experience and previous education/training appropriately and effectively; does not apply rules or strategies blindly; assesses costs and benefits of alternative solutions and makes timely decisions even with incomplete information.
LOW	**MODERATE**	**HIGH**
1 2	**3 4 5**	**6 7**

19. Information Management		
How effectively does this soldier monitor, interpret, and redistribute information received from multiple sources (especially in a digitized environment)?		
Easily experiences information overload; has trouble monitoring and interpreting multiple information sources; is unable to cope with a digitized environment; is inefficient or unable to process information and prepare it for redistribution so that it is useable by others.	Usually can handle a fair amount of information effectively; often able to effectively monitor multiple information sources, but can become overwhelmed by the speed of communication provided by digitized equipment; is able to process information and redistribute it for use by others, but fails to effectively combine or exclude information.	Can monitor, interpret, and redistribute large amounts of information received from multiple sources, especially in digitized environments; processes information effectively so that it is optimally useful to others; does not readily experience information overload.
LOW	**MODERATE**	**HIGH**
1 2	**3 4 5**	**6 7**

Section II: Overall Effectiveness

Please read the description below of overall soldier effectiveness and then rate how effective each soldier is by marking the appropriate number.

Overall Effectiveness How effectively does this soldier perform overall?		
Performs poorly in important effectiveness areas; does not meet standards for soldier performance compared to peers at same experience level.	Performs adequately in important effectiveness areas; meets standards and expectations for soldier performance compared to peers at same experience level.	Performs excellently in all or almost all effectiveness areas; exceeds standards and expectations for soldier performance compared to peers at same experience level.

Section III: Senior NCO Potential

On this rating, evaluate each soldier on his or her potential effectiveness as a senior NCO (E-7 to E-9). At this point, you are not to rate on the basis of present performance and effectiveness, but instead, indicate how well each soldier is likely to perm as a senior NCO in his or her MOS (assume each will have an opportunity to be a senior NCO). Thus, the "overall effectiveness" rating you completed in Section II and this rating of senior NCO potential may not necessarily agree closely.

Senior NCO Potential Which of the following best describes each soldier's senior NCO potential?		
Would likely be a bottom-level performer as a senior NCO.	Would likely be an adequate performer as a senior NCO.	Would likely be a top-level performer as a senior NCO.

Appendix B
Expected Future Performance Rating Scales

Expected Performance Under Future Army Conditions

Instructions

In this booklet, you will read several scenarios that describe some of the major changes predicted to occur in the future Army. After you read each scenario please rate how effectively you would expect each soldier to meet those future NCO requirements. Note that actual future Army conditions may differ from these scenarios.

Use the separately provided scannable sheet to record your ratings.

Scenario #1: Increased Requirements for Self-Direction and Self-Management

The predicted changes in missions, technology, structure, and tactics will require that NCOs have a greater ability to guide their own professional development and manage their personal affairs (e.g., family concerns and financial matters). Obviously, increasing mission diversity and frequency will be disruptive. For example, frequent deployments away from U.S. home bases will require a strong ability to manage personal matters effectively. In addition, the restructuring of the Army into smaller, more independent units will require that NCOs have a greater ability to take initiative in their actions and make their own decisions without direct supervision. Finally, due to greater technological change and more frequent changes in missions, there is an expectation that individual NCOs will need to assume more and more responsibility for their own training. That is, they will be required to identify their own training needs and to seek out training experiences that meet these needs. They will need to evaluate their own training accomplishments and take corrective steps if necessary.

1. How effectively would you expect the soldier to meet these future NCO requirements?

Not likely to meet the NCO demands described under these conditions.		Likely to be generally successful, but will struggle to meet the NCO demands described under these conditions.			Likely to successfully meet or exceed NCO demands described under these conditions.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

Scenario #2: Use of Computers, Computerized Equipment, and Digitized Operations

The digitization of the Army that started in the mid-1990s will increase and become more widespread by 2010. Commercial applications of personal computers (PCs), laptops, and small hand-held devices will become the standard means for communicating and relaying information for all soldiers, in all jobs, at all levels. Specialized military applications of computers will become more widespread and will be found on all tactical vehicles and weapons systems. Voice recognition will provide essentially hands-free operation for crewmembers. Individualized applications, available to dismounted soldiers in a variety of roles, will provide automated links for information flow in tactical settings. In addition, a tactical Internet will make it possible for operators to link to each other at all levels and locations in real time. Automation will have a serious impact on the logistical and service support functions of the Army in that most aspects of supply, maintenance, and transport will use some form of computerized system. These will start with the user of the service or supply and be linked upwards to the depot level and beyond.

While much of the focus will be on computer hardware, the truly significant advancements in technology will involve the development of specialized software. These programs will cover a variety of functions such as land navigation, orders preparation, after action analysis, and information sorting and processing. This specialized software could change how soldiers function at all levels. The Army will likely be able to automate many of the current manual functions, giving greater skills and abilities to more individuals. At the same time, specialized software will require specialized input and manipulation.

Computerization and automation will not be foolproof. System failures, clutter, jamming, hacking, interceptions, and false information are all risks that come with the use of computer-based communications. The need for back-up manual knowledge, alternate procedures, fail-safe checks, and trouble-shooting skills will place increased demands on soldier knowledge and performance. NCOs and officers will need to be able to oversee and monitor systems used by lower-level operators and implementers. In all, increased computerization will bring more, rather than less, complex demands on the NCO.

2. How effectively would you expect this soldier to meet these future NCO requirements?

Not likely to meet the NCO demands described under these conditions.		Likely to be generally successful, but will struggle to meet the NCO demands described under these conditions.			Likely to successfully meet or exceed NCO demands described under these conditions.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

Scenario #3: Increased Scope of Technical Skill Requirements

The future Army will be based on a combination of advanced weapons systems, various levels of information systems, and sophisticated communications. Organizationally, a significant part of the Army is intended to contain small, flexible battle force teams. These teams will be highly trained with a mixing of roles across ranks and with all team members cross-trained in each others' skills. The existing structure of a large number of specialized MOS likely will be replaced by a system in which NCOs are classified into broad areas of job abilities based primarily on types of units or echelons of employment. NCOs in battle forces will be expected to employ a full array of organic and supporting fires, maneuver and transportation, intelligence gathering facilities, engineering methods, data communications, and protective measures. Logistics, including supply, maintenance and repair, and field medical and evacuation will become organic requirements of the battle force. The NCO of the future will have almost unlimited access to information sources for diagnoses and step-by-step procedures, but actual performance will still have to be learned and practiced. The end result will be an increase in the technical requirements for future NCOs, probably doubling or tripling the number of skill tasks associated with today's NCOs.

3. How effectively would you expect this soldier to meet these future NCO requirements?

Not likely to meet the NCO demands described under these conditions.		Likely to be generally successful, but will struggle to meet the NCO demands described under these conditions.			Likely to successfully meet or exceed NCO demands described under these conditions.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

Scenario #4: Increased Requirements for Broader Leadership Skills at Lower Levels

Over the next 20 years, broader leadership skills will be a critical requirement of the NCO. Units the size of current platoons and companies will be the focal points of operations. Combat support and combat service support organizations will be even smaller with only 1 to 5 person cells providing specialized assistance. It will be common for units to be widely scattered and, while communication and information linkage will increase, there will be less physical contact between units of all sizes. In many situations the chain of command will be temporary and will be through information linkages rather than established relationships. Furthermore, because many missions will be situation specific, NCOs will not be able to rely as much on past experiences when making decisions in new situations.

As a result, many of the requirements for leadership, decision making, initiative, responsibility, and accountability that are today thought of as company-grade and junior officer requirements will become the domains of the E7 and E6. In turn, the level of leadership, authority, and responsibility that is currently associated with platoon sergeants, staff shift supervisors, detachment, and shop supervisors will migrate down to the E5 and E4 levels. Although at some point, future NCOs will be able to access automated decision matrices or artificial intelligence to assist them with their leadership decisions, they will have many requirements similar to what leaders have always faced – unpredicted situations, human interactions and stresses, system malfunctions, and time pressures. The difference will be that these requirements, and their consequences, will be experienced in a greater degree and at lower ranks by future NCOs.

4. How effectively would you expect this soldier to meet these future NCO requirements?

Not likely to meet the NCO demands described under these conditions.		Likely to be generally successful, but will struggle to meet the NCO demands described under these conditions.			Likely to successfully meet or exceed NCO demands described under these conditions.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

Scenario #5: Need to Manage Multiple Operational Functions and Deal with the Inter-relatedness of Units

The future Army will have a less rigid organizational structure, more mission type operations that have multiple purposes (e.g., mixed peace making/peacekeeping), more independent operations at lower levels, and increased low-level lethality. It will still employ the engagement systems of maneuver; fire support; information dominance; reconnaissance, surveillance, and intelligence; mobility and survivability; and air defense along with the integrating systems of command and control and combat service support. However, as technology and information flow improves, these will be planned for, integrated, and executed at lower and lower levels. With more capabilities at lower levels and operating under mission-type orders, NCOs will have more flexibility in the courses of actions available to them in any given situation. Along with this will come a requirement to be more aware of how one's own actions affect the total environment in which the NCO is operating. Impacts on other units, higher headquarters missions, civilian populations, strategic goals, and fratricide possibilities must be weighed by individual NCOs into any course of action they are contemplating. The ability to predict the effects of an activity onto others within the battlespace will become a crucial element of NCO-led operations. The boundaries of these operations will not be limited to what they can see or even by physical limits. NCOs must be able to operate by projecting the effects of their decisions in many directions and levels simultaneously. Although these requirements will be accompanied by improvements in technology and decision software, the timing and control of the use of available systems will remain very much a human element.

5. How effectively would you expect this soldier to meet these future NCO requirements?

Not likely to meet the NCO demands described under these conditions.		Likely to be generally successful, but will struggle to meet the NCO demands described under these conditions.			Likely to successfully meet or exceed NCO demands described under these conditions.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

Scenario #5: Need to Manage Multiple Operational Functions and Deal with the Inter-relatedness of Units

The future Army will have a less rigid organizational structure, more mission type operations that have multiple purposes (e.g., mixed peace making/peacekeeping), more independent operations at lower levels, and increased low-level lethality. It will still employ the engagement systems of maneuver; fire support; information dominance; reconnaissance, surveillance, and intelligence; mobility and survivability; and air defense along with the integrating systems of command and control and combat service support. However, as technology and information flow improves, these will be planned for, integrated, and executed at lower and lower levels. With more capabilities at lower levels and operating under mission-type orders, NCOs will have more flexibility in the courses of actions available to them in any given situation. Along with this will come a requirement to be more aware of how one's own actions affect the total environment in which the NCO is operating. Impacts on other units, higher headquarters missions, civilian populations, strategic goals, and fratricide possibilities must be weighed by individual NCOs into any course of action they are contemplating. The ability to predict the effects of an activity onto others within the battlespace will become a crucial element of NCO-led operations. The boundaries of these operations will not be limited to what they can see or even by physical limits. NCOs must be able to operate by projecting the effects of their decisions in many directions and levels simultaneously. Although these requirements will be accompanied by improvements in technology and decision software, the timing and control of the use of available systems will remain very much a human element.

5. How effectively would you expect this soldier to meet these future NCO requirements?

Not likely to meet the NCO demands described under these conditions.		Likely to be generally successful, but will struggle to meet the NCO demands described under these conditions.			Likely to successfully meet or exceed NCO demands described under these conditions.	
LOW		**MODERATE**			**HIGH**	
1	**2**	**3**	**4**	**5**	**6**	**7**

Appendix C
Conditional Means and Effect Sizes

Why Calculate Conditional Means and Effect Sizes?

As mentioned in Chapter 1, the focus of this project is on the semi-centralized NCO promotion system, covering promotions from grade E4 to E5 and from grade E5 to E6. In this system, promotion decisions are made within military occupational specialty (MOS). For example, E5 Military Police (MOS 95B) compete only with other E5 95Bs for promotion to the next pay grade. Therefore, the most useful unit of analysis for examining subgroup differences would be within MOS. However, this effort's sample sizes do not support the consideration of such differences at the MOS level. Therefore, we present subgroup differences (i.e., gender, race, and career management field [CMF] cluster) at a more aggregated level of analysis.

One disadvantage of this approach is that effects that seem to be due to one type of subgroup difference might be due to another. For example, Tables 4.6 and 4.7 in Chapter 4 present statistics for SimPPW Civilian Education scores broken down by subgroup (pay grade, race, gender, and CMF cluster). Table 4.6 shows that among E5 soldiers, the raw mean SimPPW Civilian Education score was 0.52 standard deviation higher for women than for men. However, we know that a substantial portion of the men in this study were in male-only combat MOS, and from anecdotal discussions with soldiers we learned that individuals in combat MOS report less opportunity to pursue civilian education than soldiers in other MOS. These anecdotal discussions were supported by the results shown in Table 4.7; for E5 soldiers, Combat Operations was the CMF cluster with the lowest raw mean SimPPW Civilian Education score. This means that the substantial difference in raw mean scores on this variable, favoring women by 0.52 standard deviation, might have little to do with male-female differences within any particular MOS; rather, it might be because a substantial number of the men were in combat MOS.

A potential solution to this problem, given our low sample sizes for most MOS, is to calculate conditional means and effect sizes. They offer the benefit of reflecting estimated differences between subgroups while holding other grouping variables constant. For example, comparing the conditional means of gender removes differences between males and females that are due to differences in composition of the two samples in terms of race, pay grade, and CMF cluster. For example, Table 4.6 shows that E5 women had a conditional mean SimPPW Civilian Education score only 0.15 standard deviation higher than men. The idea is that after holding other subgroup differences constant (e.g., CMF cluster), the mean difference on the SimPPW Civilian Education score, favoring women, was substantially less.

Finally, it is notable that the raw male and female soldier means on SimPPW Civilian Education are statistically different (effect size = 0.52; $p < .001$), but the conditional means are not significantly different (effect size = 0.15; $p = .428$). This means that the significant difference in the raw means was not due to differences in gender, but differences in other variables (e.g., race or CMF).

Method

Conditional means differ from raw means in that conditional means are the unweighted means of the lower-level cell means. When computing raw means, the lower-level cell means are a function of the cell (i.e., group) sample sizes. To demonstrate the difference, consider the following fictitious data:

Soldier #	Gender	Race	Score
1	M	W	5
2	M	W	7
3	M	B	4
4	F	W	3
5	F	B	4
6	F	B	5
7	F	B	6

From this table, we calculate the following cell means:

N	Gender	Race	Mean
2	M	W	6
1	M	B	4
1	F	W	3
3	F	B	5

To calculate the raw mean for each higher-order effect (gender or race), the numerator is the sum of the individual scores and the denominator is the number of individuals. Thus, for gender:

n	Gender	Raw Mean
3	M	(5+7+4)/3 = 5.3
4	F	(3+4+5+6)/4 = 4.5

and for race:

n	Race	Raw Mean
3	W	(5+7+3)/3 = 5.0
4	B	(4+4+5+6)/4 = 4.75

To calculate the conditional mean for the higher-order effects, the numerator is the sum of the lower-level cell means and the denominator is the number of cell means (i.e., number of groups in the higher-order effect). Thus, for gender:

n	Gender	Raw Mean
3	M	(6+4)/2 = 5.0
4	F	(3+5)/2 = 4.0

and for race:

n	Race	Raw Mean
3	W	(6+3)/2 = 4.5
4	B	(4+5)/2 = 4.5

As can be observed, the raw male mean is 5.3; however, after holding difference due to race constant, the conditional male mean is 5.0. Likewise, the raw white mean is higher than the raw black mean; however, after holding differences due to gender constant, the conditional white and black means are equal.

Female-male and black-white conditional effect sizes were calculated by taking the conditional mean of the non-referent group minus the conditional mean of the referent group, and dividing the resulting quantity by the pooled standard deviation for the referent group (within each pay grade). This pooled standard deviation was calculated by pooling the standard deviation associated with each subgroup combination for the referent group of interest. For example, the standard deviation underlying the conditional effect size comparing means of female and male E5 soldiers on a particular score was formed by pooling 12 standard deviations (one standard deviation across male E5 soldiers for each CMF cluster-by-race combination).

CMF cluster conditional effect sizes were calculated by taking the conditional mean of the higher numbered CMF cluster minus the conditional mean of the lower numbered CMF cluster and dividing the resulting quantity by the *overall* pooled standard deviation (within each pay grade). This *overall* pooled standard deviation was calculated by pooling the standard deviation associated with each subgroup combination for the pay grade of interest. For example, the standard deviation underlying the conditional effect size comparing means of E5 soldiers in the Administrative and Intelligence CMF clusters on a particular score was formed by pooling 24 standard deviations (one standard deviation across all E5 soldiers for each CMF cluster-by-race-by-gender combination).

Appendix D
Personnel File Form-21

Personnel File Form-21

MARKING INSTRUCTIONS

- Use a No. 2 pencil only.
- Do not use ink, ballpoint, or felt tip pens.
- Make solid marks that fill the response completely.
- Erase cleanly any marks you wish to change.
- Make no stray marks on this form.

CORRECT: ● INCORRECT: ✓ ✗ ○ ◒

ID Number

◆ Awards/Commendations

1. Mark the awards and decorations listed below that you have received. If you have received any awards or decorations not listed, mark "other" and specify the name of the award or decoration.

- ○ Soldier's Medal or higher award
- ○ Bronze Star Medal (Valor or Merit)
- ○ Defense Meritorious Service Medal
- ○ Meritorious Service Medal
- ○ Air Medal (Valor or Merit)
- ○ Joint Service Commendation Medal
- ○ Joint Achievement Medal
- ○ Purple Heart
- ○ Combat Infantry Badge
- ○ Combat Field Medical Badge
- ○ Expert Infantry Badge
- ○ Expert Field Medical Badge
- ○ Basic Parachutist Badge
- ○ Senior Parachutist Badge
- ○ Master Parachutist Badge
- ○ Divers Badge
- ○ Explosive Ordnance Disposal Badge
- ○ Pathfinder Badge
- ○ Aircraft Crewman Badge
- ○ Nuclear Reactor Operator Badge
- ○ Ranger Tab
- ○ Special Forces Tab
- ○ Driver and Mechanic Badge
- ○ Air Assault Badge
- ○ Drill Sergeant Identification Badge
- ○ US Army Recruiter Badge
- ○ Campaign Star (Battle Star)
- ○ Equivalent awards and decorations earned in other US uniformed services
- ○ Army Reserve Components Acheivement Medal
- ○ Southwest Asia Medal
- ○ Other _____
- ○ Other _____

If you received any of the following medals, indicate how many.	1	2	3 or more
Army Commendation Medal (Valor or Merit)	○	○	○
Army Achievement Medal	○	○	○
Good Conduct Medal	○	○	○

Military Academic Achievement

- ○ Distinguished Honor Graduate
- ○ Distinguished Leadership Award
- ○ Commandant's List

Military Board Achievement

- ○ Soldier/NCO of the Quarter - Brigade Level
- ○ Soldier/NCO of the Year - Brigade Level
- ○ Soldier/NCO of the Quarter - Installation/Division Level
- ○ Soldier/NCO of the Year - Installation/Division Level
- ○ Soldier/NCO of the Year - MACOM Level

2. How many Memoranda/ Letters of Appreciation, Commendation, Achievement have you received...

Write the number in the boxes.

Then, fill in the matching circle below each box.

3. How many Certificates of Achievement have you received...

◆ Military Education

4. Indicate courses listed below that you have successfully completed. Do not include BT, OSUT, or AIT.

- ◯ PLDC
- ◯ Airborne School
- ◯ BNCOC - If yes, how many weeks? ⟶
- ◯ NBC School
- ◯ Ranger School
- ◯ Air Assault School
- ◯ Special Forces Qualification Course
- ◯ Any other course of at least 40 hours duration - If yes, how many? ⟶
- ◯ Military correspondence course credit hours - If yes, how many? ⟶
- ◯ EMT Basic Certification
- ◯ EMT Intermediate Certification
- ◯ EMT Paramedic Certification

◆ Civilian Education

5. List the total number of semester hours you have earned since you have been on active duty.

a. Career/ Trade School

b. Vo Tech

c. College

Of the semester hours you have earned since you have been on active duty, indicate how many were paid for through the Army's Tuition Assistance Program.

a. Career/ Trade School

b. Vo Tech

c. College

6. Have you earned a civilian college degree since you have been on active duty?

- ◯ Yes - If yes, indicate the type of degree(s)
  - ◯ Associates
  - ◯ Bachelors
  - ◯ Masters
  - ◯ Other_____
- ◯ No

If you answered yes to Question 6, indicate when you started to work on your degree and when you completed it.

Started		Finished	
Mo.	Yr.	Mo.	Yr.

◆ Disciplinary Action

7. How many Articles 15 have you received. . .

8. How many Flag Actions (i.e., suspension of favorable personnel action) have you received. . .

◆ Test Scores

9. What was your last Physical Readiness Test score? (score ranges from 0-300)

10. What was your last Weapon Qualification?
- ○ Unqualified
- ○ Marksman (MKM)
- ○ Sharpshooter (SPS)
- ○ Expert (EXP)

11. Have you retaken the ASVAB since your initial enlistment screening?
- ○ Yes - If yes, how many times have you retaken the ASVAB/AFCT exam? →

- ○ No

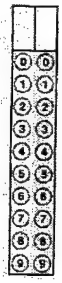
12. What is your current General Technical (GT) score of record?

◆ ACES Participation

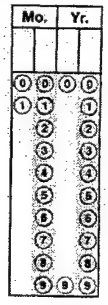
This section asks about your participation in programs sponsored by the Army Continuing Education System (ACES).

13. How many MOS Improvement/ Soldier (Unit) Training Courses sponsored by Army Education have you successfully completed?

14. a. How many Army Education NCO Leadership Development Courses did you successfully complete prior to being promoted to your current grade?



b. When did you complete the last NCO Leadership Development Course prior to being promoted to your current grade?



- ○ Not applicable

next page.

15. **To what extent have Army Education programs such as Tuition Assistance, college/vocational-technical courses, NCO Leadership Development Courses, and MOS Improvement Courses improved your competence to perform at the next higher grade level?**

- ○ Does not apply; I have not participated in any Army Education programs.
- ○ Army Education programs have not improved my competence.
- ○ Army Education programs have slightly improved my competence.
- ○ Army Education programs have somewhat improved my competence.
- ○ Army Education programs have greatly improved my competence.

16. **To what extent have Army Education programs enhanced your performance as a soldier?**

- ○ Does not apply; I have not participated in any Army Education programs.
- ○ Army Education programs have not enhanced my performance.
- ○ Army Education programs have slightly enhanced my performance.
- ○ Army Education programs have somewhat enhanced my performance.
- ○ Army Education programs have greatly enhanced my performance.

PLEASE DO NOT WRITE IN THIS AREA

Appendix E
Experience and Activities Records

Experience & Activities Record

This form lists a variety of experiences, activities, or assignments that some soldiers have had. Please respond to each item based on your experience.

Experiences and Activities	Frequency In the last 2 years, how often have you performed each activity?					
	Never	A few times a year	About once a month	A few times a month	A few times a week	Daily
Computer Related Activities						
1. Used a PC, Mac, or laptop.	☐	☐	☐	☐	☐	☐
2. Communicated using e-mail.	☐	☐	☐	☐	☐	☐
3. Used the Internet for job or training requirements.	☐	☐	☐	☐	☐	☐
4. Used the Windows NT operating system.	☐	☐	☐	☐	☐	☐
5. Operated an Army-specific computer system (e.g., IVIS, ASAS, FBCB2, AFATDS).	☐	☐	☐	☐	☐	☐
6. Troubleshooted a computer system malfunction.	☐	☐	☐	☐	☐	☐
7. Used Windows Office programs to do job tasks (e.g., Word®, Access®, Excel®, PowerPoint®).	☐	☐	☐	☐	☐	☐
8. Trained or assigned as an instructor/operator (I/O) on any computer based simulator (e.g., COFT, BBS, CBS, SIMNET, Janus).	☐	☐	☐	☐	☐	☐
Leadership/Supervisory						
9. Assigned to duty position with a responsibility for supervising 2 or more soldiers.	☐	☐	☐	☐	☐	☐
10. Provided performance feedback to subordinates.	☐	☐	☐	☐	☐	☐
11. Established goals or other incentives to motivate subordinates.	☐	☐	☐	☐	☐	☐
12. Corrected unacceptable conduct of a subordinate.	☐	☐	☐	☐	☐	☐
13. Trained other soldiers in a task or a procedure.	☐	☐	☐	☐	☐	☐
14. Conducted formal inspection of subordinates' completed work.	☐	☐	☐	☐	☐	☐
15. Counseled subordinates regarding career planning.	☐	☐	☐	☐	☐	☐
16. Counseled subordinates with disciplinary problems.	☐	☐	☐	☐	☐	☐
17. Served as a member of a unit advisory council or committee.	☐	☐	☐	☐	☐	☐
18. Applied and supervised all 8 steps of troop leading procedures (TLP).	☐	☐	☐	☐	☐	☐

Experiences and Activities (continued)

	Never	A few times a year	About once a month	A few times a month	A few times a week	Daily
Additional Duties						
19. Volunteered for additional duties/assignments.	☐	☐	☐	☐	☐	☐
20. Requested additional training opportunities.	☐	☐	☐	☐	☐	☐

Assignments and Positions

	Duration How much time have you spent in each of the following?				
	Never	Less than 6 Months	6 months to a year	1 year to 2 years	More than 2 years
Duration of Experiences					
21. Total time spent in duty position one grade higher than actual grade.	☐	☐	☐	☐	☐
22. Total time spent in a leadership or supervisory position.	☐	☐	☐	☐	☐
23. Total time spent in MTOE slot assignment.	☐	☐	☐	☐	☐
24. Total time in a unit specialty assignment (e.g., Commander's or First Sergeant's driver, Assistant Training NCO, NBC, Unit Lifesaver).	☐	☐	☐	☐	☐

Training and Duties	Frequency How many times have you done each of the following? Never	Once	Twice or more
Formal Training/Assignments			
25. Participated in CTC/NTC/JRTC rotation or FTX over 30 days.	☐	☐	☐
26. Deployed on combat mission.	☐	☐	☐
27. Deployed on peace-keeping mission.	☐	☐	☐
28. Prepared a lesson plan.	☐	☐	☐
29. Led a PT class.	☐	☐	☐
30. Taught a platform class to 5 or more people.	☐	☐	☐
31. Served as an assistant instructor in a class of 10 or more people.	☐	☐	☐
32. Been part of a crew to perform Table VIII, Table XII, or TCPC.	☐	☐	☐
33. Participated as a team leader or above in a live fire exercise (LFX).	☐	☐	☐
34. Conducted primary marksmanship instruction (PMI).	☐	☐	☐
Communications			
35. Received and implemented a written operations order.	☐	☐	☐
36. Issued a 5 paragraph oral operations order.	☐	☐	☐
37. Prepared and submitted a written report of recognition for a subordinate.	☐	☐	☐
38. Prepared and conducted a briefing for 2 or more officer, senior NCO, or civilian personnel.	☐	☐	☐
39. Prepared a written plan/schedule of future subordinate activities covering 5 days or more.	☐	☐	☐
40. Prepared a written counseling statement.	☐	☐	☐
Inspections, Drills and Ceremonies, Official Duties			
41. Led/commanded soldiers in drill and ceremony activities.	☐	☐	☐
42. Conducted an inspection in ranks or standby.	☐	☐	☐
43. Performed as Color Guard.	☐	☐	☐
44. Acted as assistant commander at funeral detail or other public ceremony.	☐	☐	☐
45. Served as a VIP escort.	☐	☐	☐
46. Appeared before a Soldier of the Month (or equivalent) Board.	☐	☐	☐

Appendix F
E4 Predictor Score Correlations

Table F.1. Raw Correlations among Predictor Scores for E4 Soldiers

Predictors	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1 ASVAB GT																									
2 SJT Composite	**.18**																								
3 Interview Composite	.06	**.11**																							
4 Simulated PPW Composite	.02	.00	**.11**																						
5 PPW Awards	-.01	-.04	.07	**.65**																					
6 PPW Military Education	.05	**.10**	**.11**	**.59**	**.17**																				
7 PPW Civilian Education	.01	.04	.05	**.39**	.02	.06																			
8 PPW Military Training	.04	**-.08**	.03	**.55**	**.12**	**.12**	.00																		
9 ExAct Computer Experience	**.14**	**.13**	.06	**.09**	.00	**.13**	.07	.01																	
10 ExAct Supervisory Experience	-.06	.01	**.14**	**.41**	**.31**	**.20**	**.08**	**.32**	.06																
11 ExAct General Experience	.06	**.09**	**.19**	**.48**	**.40**	**.23**	.06	**.34**	**.20**	**.66**															
12 AIM Dependability	.01	**.33**	**.17**	.00	**-.10**	**.14**	**.14**	**-.14**	**.12**	-.06	-.03														
13 AIM Adjustment	.07	**.29**	**.11**	.07	-.02	.06	.06	.07	**.08**	.07	.05														
14 AIM Work Orientation	.05	**.31**	**.29**	**.11**	-.03	**.12**	**.08**	**.09**	**.10**	**.14**	**.18**	**.42**	**.34**												
15 AIM Agreeableness	.03	**.31**	**.13**	-.05	**-.11**	.05	.05	**-.09**	.06	-.07	-.06	**.55**	**.47**	**.39**											
16 AIM Physical Conditioning	.02	**.20**	**.13**	**.12**	-.07	.04	**.13**	**.20**	.00	.03	.02	**.32**	**.32**	**.43**	**.28**										
17 AIM Leadership	**.13**	**.26**	**.32**	**.12**	.04	.07	**.08**	**.10**	**.11**	**.22**	**.26**	**.25**	**.40**	**.60**	**.21**	**.10**									
18 BIQ Hostility to Authority	**-.17**	**-.26**	-.08	**.08**	**.12**	-.06	.04	.07	-.07	**.08**	.04	**-.45**	**-.26**	**-.30**	**-.35**	**-.21**	**-.16**								
19 BIQ Manipulativeness	**-.14**	**-.32**	**-.18**	-.01	.05	-.05	**-.11**	**.09**	-.07	.07	-.01	**-.44**	**-.33**	**-.38**	**-.38**	**-.21**	**-.26**	**.61**							
20 BIQ Social Perceptiveness	.07	**.18**	**.15**	.03	-.05	-.01	**.13**	.04	**.15**	.06	**.11**	**.14**	**.25**	**.32**	**.14**	**.14**	**.38**	.04	**-.14**						
21 BIQ Social Maturity	**.16**	**.31**	**.10**	-.02	-.06	**.11**	.07	**-.17**	**.12**	**-.10**	-.05	**.54**	**.16**	**.23**	**.33**	**.18**	**.09**	**-.67**	**-.66**	-.02					
22 BIQ Tolerance for Ambiguity	**.16**	**.24**	**.19**	**.10**	**.09**	.03	.04	.06	**.20**	.01	**.17**	**.11**	**.25**	**.27**	**.19**	**.11**	**.34**	**-.29**	**-.36**	**.23**	**.24**				
23 BIQ Openness	**.11**	**.12**	**.12**	.08	.05	-.01	**.08**	.04	**.20**	**.09**	**.17**	.07	**.18**	**.19**	**.17**	**.10**	**.25**	.04	-.06	**.46**	-.03	**.32**			
24 BIQ Leadership	**.11**	**.17**	**.25**	**.16**	.03	.08	**.12**	**.16**	**.18**	**.28**	**.30**	**.09**	**.31**	**.41**	.08	**.13**	**.60**	.04	**-.09**	**.66**	-.06	**.27**	**.42**		
25 BIQ Interpersonal Skill	**.18**	**.35**	**.16**	-.03	-.04	.03	-.04	-.03	**.14**	-.02	.05	**.37**	**.37**	**.32**	**.38**	**.13**	**.31**	**-.54**	**-.52**	**.24**	**.40**	**.36**	**.08**	**.20**	

Note. n = 290–448. Statistically significant correlations are bolded, $p < .05$ (one-tailed).